Die Humankybernetik (Anthropokybernetik) umfaßt alle jene Wissenschaftszweige, welche nach dem Vorbild der neuzeitlichen Naturwisseschaft versuchen, Gegenstände, die bisher ausschließlich mit geisteswissenschaftlichen Methoden bearbeitet wurden, auf Modelle abzubilden und mathematisch zu analysieren. Zu den Zweigen der Humankybernetik gehören vor allem die Informationspsychologie (einschließlich der Kognitionsforschung, der Theorie über „künstliche Intelligenz" und der modellierenden Psychopathometrie und Geriatrie), die Informationsästhetik und die kybernetische Pädagogik, aber auch die Sprachkybernetik (einschließlich der Textstatistik, der mathematischen Linguistik und der konstruktiven Interlinguistik) sowie die Wirtschafts-, Sozial und Rechtskybernetik. - Neben diesem ihrem hauptsächlichen Themenbereich pflegen die GrKG/Humankybernetik durch gelegentliche Übersichtsbeiträge und interdisziplinär interessierende Originalarbeiten auch die drei anderen Bereiche der kybernetischen Wissenschaft: die Biokybernetik, die Ingenieurkybernetik und die Allgemeine Kybernetik (Strukturtheorie informationeller Gegenstände). Nicht zuletzt wird auch metakybernetischen Themen Raum gegeben: nicht nur der Philosophie und Geschichte der Kybernetik, sondern auch der auf kybernetische Inhalte bezogenen Pädagogik und Literaturwissenschaft. -

La prihoma kibernetiko (antropokibernetiko) inkluzivas ĉiujn tiajn sciencobranĉojn, kiuj imitante la novepokan natursciencon, klopodas bildigi per modeloj kaj analizi matematike objektojn ĝis nun pritraktitajn ekskluzive per kulturscienca j metodoj. Apartenas al la branĉaro de la antropokibernetiko ĉefe la kibernetika psikologio (inkluzive la ekkon-esploron, la teoriojn pri „artefarita intelekto" kaj la modeligajn psikopatometrion kaj geriatrion), la kibernetika estetiko kaj la kibernetika pedagogio, sed ankaŭ la lingvokibernetiko (inkluzive la tekststatisikon, la matematikan lingvistikon kaj la konstruan interlingvistikon) same kiel la kibernetika ekonomio, la socikibernetiko kaj la jurkibernetiko. - Krom tiu ĉi sia ĉefa temaro per superrigardaj artikoloj kaj interfake interesigaj originalaj laboraĵoj GrKG/HUMANKYBERNETIK flegas okaze ankaŭ la tri aliajn kampojn de la kibernetika scienco: la biokibernetikon, la inĝenierkibernetikon kaj la ĝeneralan kibernetikon (strukturteorion de informecaj objektoj). Ne lastavice trovas lokon ankaŭ metakibernetikaj temoj: ne nur la filozofio kaj historio de la kibernetiko, sed ankaŭ la pedagogio kaj literaturscienco de kibernetikaj sciaĵoj. -

Cybernetics of Social Systems comprises all those branches of science which apply mathematical models and methods of analysis to matters which had previously been the exclusive domain of the humanities. Above all this includes *information psychology* (including theories of cognition and 'artificial intelligence' as well as psychopathometrics and geriatrics), *aesthetics of information* and *cybernetic educational theory*, *cybernetic linguistics* (including text-statistics, mathematical linguistics and constructive interlinguistics) as well as *economic, social* and *juridical cybernetics*. - In addition to its principal areas of interest, the GrKG/HUMANKYBERNETIK offers a forum for the publication of articles of a general nature in three other fields: *biocybernetics, cybernetic engineering* and *general cybernetics* (theory of informational structure). There is also room for *metacybernetic* subjects: not just the history and philosophy of cybernetics but also cybernetic approaches to education and literature are welcome.

La cybernétique sociale contient tous le branches scientifiques, qui cherchent à imiter les sciences naturelles modernes en projetant sur des modèles et en analysant de manière mathématique des objets, qui étaient traités auparavant exclusivement par des méthodes des sciences culturelles ("idéographiques"). Parmi les branches de la cybernétique sociale il y a en premier lieu la psychologie informationelle (inclues la recherche de la cognition, les théories de l'intélligence artificielle et la psychopathométrie et gériatrie modeliste), l'esthétique informationelle et la pédagogie cybernétique, mais aussi la cybernétique linguistique (inclues la statistique de textes, la linguistique mathématique et l'interlinguistique constructive) ainsi que la cybernétique en économie, sociologie et jurisprudence. En plus de ces principaux centres d'intérêt la revue HUMANKYBERNETIK s'occupe – par quelques articles de synthèse et des travaux originaux d'intérêt interdisciplinaire – également des trois autres champs de la science cybernétique: la biocybernétique, la cybernétique de l'ingénieur et la cybernétique générale (théorie des structures des objets informationels). Une place est également accordée aux sujets métacybernétiques mineurs: la philosophie et l'histoire de la cybernétique mais aussi la pédagogie dans la mesure où elle concernent la cybernétique.

ISSN 0723–4899

Grundlagenstudien aus Kybernetik und Geisteswissenschaft

Internationale Zeitschrift für Modellierung und Mathematisierung in den Humanwissenschaften
Internacia Revuo por Modeligo kaj Matematikizo en la Homsciencoj
International Review for Modelling and Application of Mathematics in Humanities
Revue internationale pour l'application des modèles et de la mathématique en sciences humaines



HUMANKYBERNETIK

Inhalt * Enhavo * Contents * Matières

Band 24 * Heft 4/83

A. Luzzi
L. Ferranti
M. Massone
genaue Namen in 25/3, 131



Gunter Narr Verlag Tübingen

Die Grundlagenstudien aus Kybernetik und Geisteswissenschaft (GrKG/Humankybernetik) wurden 1960 durch Max Bense, Gerhard Eichhorn und Helmar Frank begründet. Sie sind z.Zt. offizielles Organ folgender wissenschaftlicher Einrichtungen:

Institut für Kybernetik Berlin e.V. (Direktor: Prof. Dr. Uwe LEHNERT, Freie Universität Berlin)

LBA - Deutsche Liga zur Bekämpfung frühzeitiger Alterserkrankungen (Präsident: Prof. Dr. med. Bernd FISCHER, Universität Heidelberg und Mannheim)

Grundlagenstudien aus Kybernetik und Geisteswissenschaft

Internationale Zeitschrift für Modellierung und Mathematisierung in den Humanwissenschaften
Internacia Revuo por Modeligo kaj Matematikizo en la Homsciencoj
International Review for Modelling and Application of Mathematics in Humanities
Revue internationale pour l'application des modèles et de la mathématique en sciences humaines



Inhalt * Enhavo * Contents * Matières

Band 24 * Heft 4/83

Gunter Narr Verlag · Stauffenbergstraße 42 · Postfach 2567 · 7400 Tübingen

Verlag und Anzeigenverwaltung | *Eldonejo kaj anonc-administrejo* | Publisher and advertisement administrator | *Edition et administration des anonnces*

Gunter Narr Verlag
Stauffenbergstraße 42, Postfach 2567, D-7400 Tübingen 1, Tel. (0049- /0-)7071 - 24156

Die Zeitschrift erscheint vierteljährlich (März, Juni, September, Dezember). Redaktionsschluß: 1. des Vormonats. - Die Bezugsdauer verlängert sich jeweils um ein Jahr, wenn bis zum 1.Dezember keine Abbestellung vorliegt. - Die Zusendung von Manuskripten (gemäß den Richtlinien auf der dritten Umschlagseite) wird an die Schriftleitung erbeten, Bestellungen und Anzeigenaufträge an den Verlag. - Z.Zt. gültige Anzeigenpreisliste: Nr. 3 vom 1.1.1982.

La revuo aperadas kvaronjare (marte, junie, septembre, decembre). Redakcia limdato: la 1-a de la antaŭa monato. - La abondaŭro plilongiĝadas je unu jaro se ne alvenas malmendo ĝis la 1-a de decembro. - Bv. sendi manuskriptojn (laŭ la direktivoj sur la tria kovrilpaĝo) al la redaktejo, mendojn kaj anoncojn al la eldonejo. - Validas momente la anoncprezlisto 3 de 1982-01-01.

This journal appears quarterly (every March, June, September and December). Editorial deadline is the 1st of the previous month. - The subscription is extended automatically for another year unless cancelled by the 1st of December. - Please send your manuscripts (fulfilling the conditions set out on the third cover page) to the editorial board, subsription orders and advertisements to the publisher. - Current prices for advertisements: List no. 3 dated 1-1-82.

La revue paraît trimestriellement (en mars, juin, septembre, décembre). Date limite pour la rédaction: le 1er du mois précédent. – L'abonnement se renouvellera automatiquement pour un an, sauf révocation reçue au plus tard le 1er décembre. – Veuillez envoyer, s.v.p., des manuscrits (suivant les indications sur la troisième page de la couverture) à l'adresse de la rédaction, des abonnements et des commandes d'annonces à celle des éditions. – Le tarif actuel en vigeur est celui des annonces du 1982-01-01.

Druck: Müller + Bass, Tübingen

ISSN 0723–4899

grkg/Humankybernetik
Band 24 · Heft 4 (1983)
Gunter Narr Verlag Tübingen

Zählungen zur Phonemhäufigkeitsverteilung des Esperanto

von Jochem SOTSCHECK, Berlin

Forschungsinstitut der Deutschen Bundespost, Außenstelle Berlin

1 Einführung

Für die Messung der Sprachverständlichkeit, einem Merkmal zur Beurteilung der Sprachübertragungsgüte in der Nachrichtentechnik, werden seit vielen Jahrzehnten als Testsprachproben neben anderen Sprachbeispielen sogenannte Esperanto-Logatome verwendet (CCIT, o.J.). Unter Logatomen versteht man hierbei bestimmte, nach vorgegebenen Bildungsvorschriften aufgestellte, sinnleere Sprachbausteine. Bei der Untersuchung der Eigenschaften dieser Testsprachproben bestand der Wunsch, deren sprachliche Eigenschaften zu analysieren und mit den Eigenschaften bestimmter Sprachen zu vergleichen.

Für diese Fragestellung sollten bestimmte Aussagen zur Phonologie der Esperanto-Logatome einerseits und zur Phonologie der damit zu vergleichenden Sprachen andererseits gegenübergestellt werden. Unter den verschiedensten phonologischen Merkmalen von Sprachen und Subsprachen sind hierzu Phoneminventar und Auftrittshäufigkeit dieser Phoneme die bedeutungsvollsten. Neben Vergleichen mit den wichtigen Verkehrssprachen Englisch und Französisch sollten die Esperanto-Logatome auch dem Esperanto selbst gegenübergestellt werden.

Nun lagen dem Verfasser weder Angaben zur Phonemhäufigkeitsverteilung des Esperanto vor, noch waren derartige Zählergebnisse nachweisbar. Deshalb wurde zunächst diese engere Fragestellung nach der Phonemhäufigkeitsverteilung des Esperanto Ziel eigener Untersuchungen. Die Ergebnisse der daraufhin durchgeführten Phonemzählungen an bestimmten Esperanto-Beispielen, die in Zusammenhang mit der oben skizzierten übergeordneten Fragestellung nur indirekt und am Rande erwähnt werden konnten (Sotscheck, 1982), sollen nun hier erstmals unter den Aspekten der Textstatistik eingehender dargestellt werden.

2 Vorgehensweise bei den Phonemzählungen

2.1 Umfang des Zählstoffes

Zunächst mußten Festlegungen über den zu untersuchenden Umfang des Zählstoffes getroffen werden. Hierzu wurde von bekannten Zählungen der Phonem- und Lauthäufigkeit anderer Sprachen ausgegangen. Entsprechende Zählergebnisse lagen vor für Deutsch (Meier, 1967), Französisch (Chavasse, 1948; Malécot, 1974) sowie Englisch

und US-Amerikanisch (Denes, 1963; Dewey, 1950; French, Carter und Koenig, 1930). Die untersuchten Sprachtexte umfaßten dabei – soweit angegeben – je Zählung zwischen 30 000 und 200 000 Sprachlaute.

Bei der Festlegung des Textumfanges der Zählstichprobe und der Anzahl der zu unterscheidenden Sprachelemente muß jeweils die beabsichtigte Verwendung der Zählergebnisse beachtet werden. So galten die angeführten Zählungen teils grundlegend linguistischen, teils speziell nachrichtentechnischen Fragestellungen. Entsprechend wurde einerseits zwischen Sprachlauten unterschieden, es wurden also sehr feine Differenzierungen vorgenommen, andererseits Zählungen nur auf unterschiedliche Phoneme erstreckt, also ausschließlich die bedeutungsunterscheidenden Sprachlaute erfaßt. Bei der Verwendung der Zählergebnisse für Aufgaben der Sprachverständlichkeitsbeurteilung in der Nachrichtentechnik ist die Unterscheidung zwischen Phonemen die zweckdienlichere, da gerade diese Sprachelemente den Wortsinn ausdrücken, der vom Gesprächspartner verstanden werden soll.

Die Größe der Textstichprobe zur Phonemzählung ist sicherlich auch von der Größe des Phoneminventars der betreffenden Sprache abhängig. Während die Sprachen Deutsch, Französisch und Englisch jede um 40 bedeutungsunterscheidende Sprachlaute besitzen, weist das Esperanto deutlich weniger Phoneme auf. Werden zwar die Diphthonge (vokalische Doppellaute) zu den eigenständigen Phonemen gerechnet, die Affrikaten (konsonantische Doppellaute aus Plosiv und Frikativ) aber, wie oft üblich, nicht – obwohl gerade sie im Esperanto durch eigene Schriftzeichen dargestellt werden –, so erhält man 29 Phoneme des Esperanto, also etwa nur dreiviertel der Menge des Phoneminventars des Deutschen. Es sollte erwähnt werden, daß beim Esperanto, wie bei anderen Sprachen auch, die Zuordnungen der Sprachlaute zu den Phoneminventaren nicht ganz einheitlich ist und vom Standpunkt des jeweiligen Autors beeinfluß wird (siehe z.B. Hölscher, 1965; Janot-Giorgetti, 1982; Wagner, 1964).

Aus den angeführten Gründen und um bei ausreichender Genauigkeit die Zählung in der kurzen zur Verfügung stehenden Zeit durchführen zu können, erschien es zulässig, den Zählstoff für die Ermittlung der Phonemhäufigkeit des Esperanto auf insgesamt 50 000 Phoneme laufenden Textes festzulegen. Der Zählstoff liegt damit in der Nähe der Untergrenze des Textumfanges der angeführten anderen Zählungen und besitzt aus diesem Grunde unter Berücksichtigong des geringeren Phoneminventars mindestens die Genauigkeit dieser Zählungen.

2.2 Auswahl des Zählstoffs

Die Textbeispiele zur Berechnung der Phonemhäufigkeitsverteilung des Esperanto sollten aus möglichst vielen und unterschiedlichen Quellen ausgewählt werden. So sollten sie ursprünglich von möglichst vielen Autoren aus unterschiedlichen Zeitepochen stammen, ferner sollten die Textbeispiele aus möglichst vielen verschiedenen Ursprungssprachen in das Esperanto übersetzt worden sein. Um auch durch die Übersetzung die Vielfalt der sprachlichen Darstellung möglichst beizubehalten, wurden zur Vermeidung von Sprachverflachungen nur solche Textbeispiele für die Zählung ausgewählt, die von Übersetzern mit der Muttersprache des jeweiligen Textautors in das Esperanto übertragen worden waren. Ausnahme war dabei der Textteil aus dem Alten Testament, das

von Zamenhof selbst, dem Schöpfer des Esperanto, aus dem Hebräischen in das Esperanto übersetzt worden war.

Es wurden als Zähltexte schließlich zehn verschiedene Literaturbeispiele mit je 5 000 Phonemen laufendem Text gewählt, die ursprünglich aus sieben verschiedenen Sprachen stammten. Die Textbeispiele stammen aus folgenden Werken:

Deutsche Texte
1. Karl Lebrecht Immermann: Der Karneval und die Somnambüle (1829)
2. Stefan Zweig (1881–1942): Die Augen des ewigen Bruders

Englischer Text
3. Gilbert Keith Chesterton: The blue cross. Aus: The innocence of Father Brown (1911)

Französische Texte
4. Guy de Maupassant (1850–1893): Boitelle. Aus: Les contes Normands
5. François Duc de La Rochefoucauld: Réflexions ou sentences et maximes morales (1665)
6. François Marie Arouet de Voltaire: Candide ou l'optimisme (1759)

Norwegischer Text
7. Knut Hamsun: Victoria (1898)

Schwedischer Text
8. August Strindberg: Die Insel der Seligen (1883)

Spanischer Text
9. Miguel de Cervantes Saavedra: Novela del curioso impertinente (Novelle von der törichten Neugier). 33. Kapitel aus: El ingenioso hidalgo Don Quijote de la Mancha (1604)

Hebräischer Text
10. Die Bibel: Genesis, Kapitel 6–8

Die in dieser Aufstellung enthaltenden Veröffentlichungsdaten oder zumindest die angegebenen Lebensdaten der Autoren weisen auf die vielfältigen Zeitepochen hin, aus denen die Literaturbeispiele entnommen wurden. Im Schrifttumsverzeichnis zu diesem Aufsatz finden sich weitere bibliographische Angaben zu den Fundstellen des Zählstoffes.

2.3 Durchführung der Zählung

Eine sprachliche Besonderheit des Esperanto liegt darin, daß jedes Phonem nur durch ein einziges verabredetes Schriftzeichen wiedergegeben wird, und umgekehrt jedes Schriftzeichen im allgemeinen nur eine einzige Lautbedeutung hat. Mehrdeutigkeiten in der Aussprache der Buchstabenreihungen, die die Wörter bilden, wie in den meisten anderen Sprachen, treten im Esperanto nicht auf.

Diese Tatsache der eindeutigen Aussprache der Schriftzeichen begünstigt eine Zählung der Auftrittshäufigkeiten der Phoneme des Esperanto in starkem Maße, weil sich die sonst sehr mühsame Transkription des Schrifttextes in einen phonetischen Text

damit erübrigt und sich die Ermittlung der Phonemhäufigkeitsverteilung – unter Beachtung einiger bestimmter Regeln – auf eine Ermittlung der Buchstabenhäufigkeitsverteilung beschränkt. Das ist aber, zumal mit Hilfe von Datenverarbeitungsanlagen, ohne allzu großen Aufwand relativ einfach auszuführen.

Die Phonem- bzw. Buchstabenhäufigkeitszählung wurde mit Hilfe eines entsprechenden Programms mit einer Datenverarbeitungsanlage durchgeführt, nachdem von je mindestens 5000 laufenden Phonemen jeder Textprobe ein Lochstreifen angefertigt worden war. Zur Darstellung einiger Phoneme bzw. Phonembündel mußten für die weitere Verarbeitung bei der Lochstreifenherstellung bestimmte Verabredungen getroffen werden (z.B. für die Wiedergabe der Buchstaben mit Accent circonflexe, für die Zählung der auch im Esperanto durch zwei Buchstaben dargestellten Diphthonge als nur ein einziges Phonem, für die Aufspaltung von einigen Einzelbuchstaben in zwei getrennte Phoneme usw.). Häufige Wiederholungen von Eigennamen sowie Eigennamen, die nicht in das Esperanto übersetzt waren, wurden nicht in die Zähltexte übernommen.

Als wichtigstes Zählergebnis wurden die Auftrittshäufigkeiten der 29 Phoneme des Esperanto in jedem, 5000 laufende Phoneme umfassenden "Volltext" der zehn Textbeispiele bestimmt. Die hierbei zwischen den Textbeispielen auftretenden Ergebnisstreuungen können möglicherweise auf autorenspezifische Unterschiede in den einzelnen Phonemhäufigkeitsverteilungen hinweisen (über die Auswertung der Zählergebnisse wird im nächsten Abschnitt berichtet).

Neben dieser Hauptzählung der Phoneme der Volltexte wurden Teilzählungen an Teiltexten aus je 1000 laufenden Phonemen vorgenommen. Die Teilzählungen sollten dazu dienen, Rückschlüsse – möglicherweise auch autorenspezifische – aus der Größe der hierbei auftretenden Streubereiche zu ziehen. Um eine Vorstellung vom Umfang dieser Teiltexte von je 1000 Phonemen Länge zu vermitteln, sei angegeben, daß diese etwa dem Textumfang einer halben Seite DIN A4 mit 1 1/2-zeiliger Schreibmaschinenschrift entsprechen. Innerhalb dieses Textumfanges sind sicherlich die individuellen Eigenschaften der verwendeten Sprache noch erkennbar.

3 *Zählergebnisse*

Einen Gesamtüberblick über die Ergebnisse der Untersuchungen zur Phonemhäufigkeitsverteilung an den zehn Textbeispielen des Esperanto, also aus insgesamt 10 x 5000 = 50 000 Phonemen, zeigt Bild 1. Hierin ist die Phonemhäufigkeit in Prozent als Funktion der 29 verschiedenen Phoneme des Esperanto dargestellt. Die Phoneme auf der Abszissenachse sind dabei nach fallender mittlerer Häufigkeit im Esperanto geordnet. Parameter in diesem Schaubild sind jeweils für jedes einzelne Phonem die Einzelergebnisse für die 10 verschiedenen Textproben Volltext. Für jedes Phonem sind durch nebeneinanderliegende Ordinatenstrecken diese Teilergebnisse aus jeder der 10 Textproben aus je 5000 Phonemen dargestellt. Die Reihenfolge der 10 einander parallelen Ordiantenstrecken für jedes Phonem entspricht der Reihenfolge 1 bis 10 der Numerierung der oben aufgeführten Werke, aus denen die Textproben entnommen worden sind. Die angegebenen Ordinatenstrecken werden durch die Extremalwerte für die Phonemhäufigkeiten begrenzt, die bei der 1000-Phoneme-Untersuchung gefunden wurden, und sie schließen den hervorgehobenen Ordinatenpunkt für den Häufigkeitsmittel-

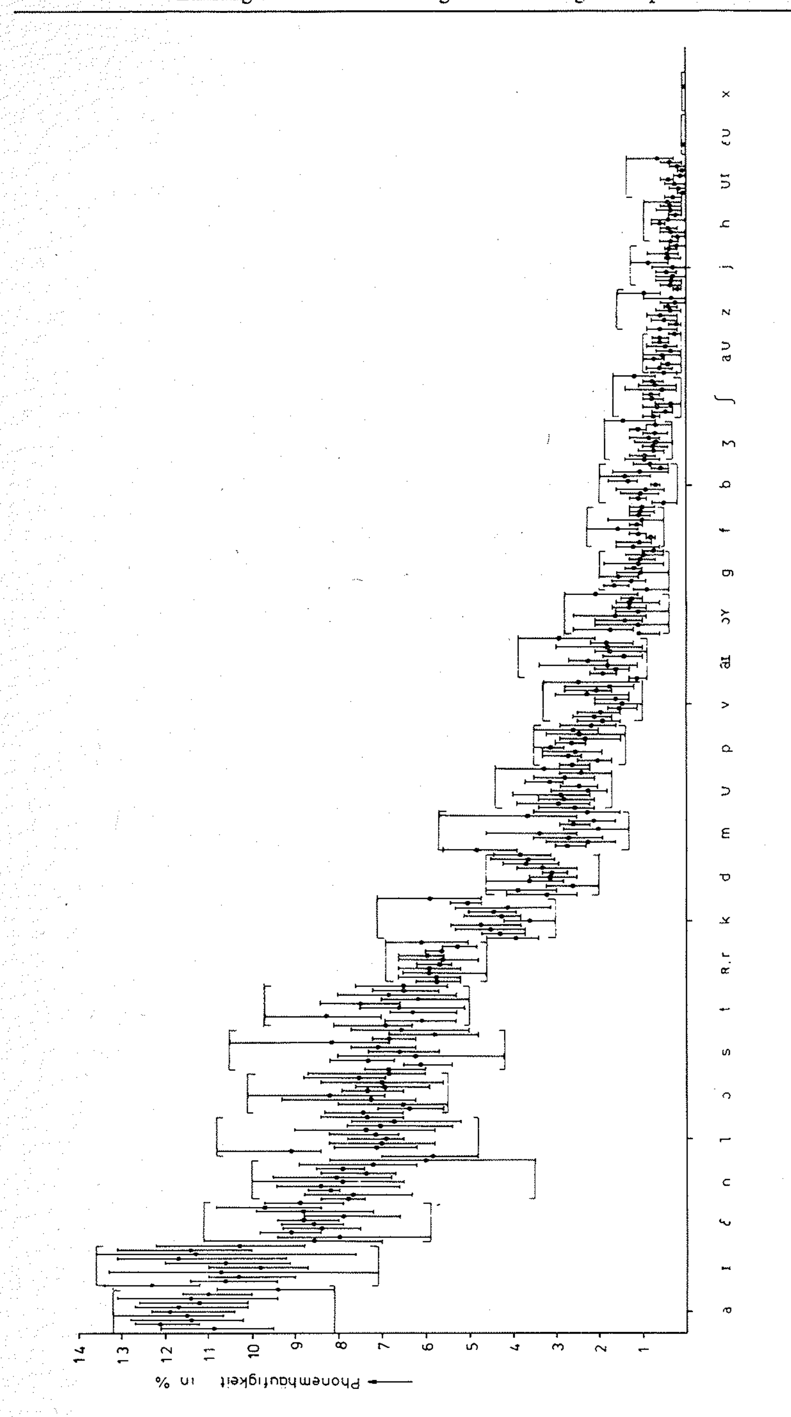


Bild 1: Phonemhäufigkeitsverteilung des Esperanto: Mittel- und Extremalwerte der untersuchten zehn Textbeispiele. Mittelwerte beziehen sich auf Volltext jedes Textbeispiels (5000 Phoneme), Extremalwerte stammen aus Zählungen an Teiltexten (1000 Phoneme). Reihenfolge der Texte gemäß Abschnitt 2.2.

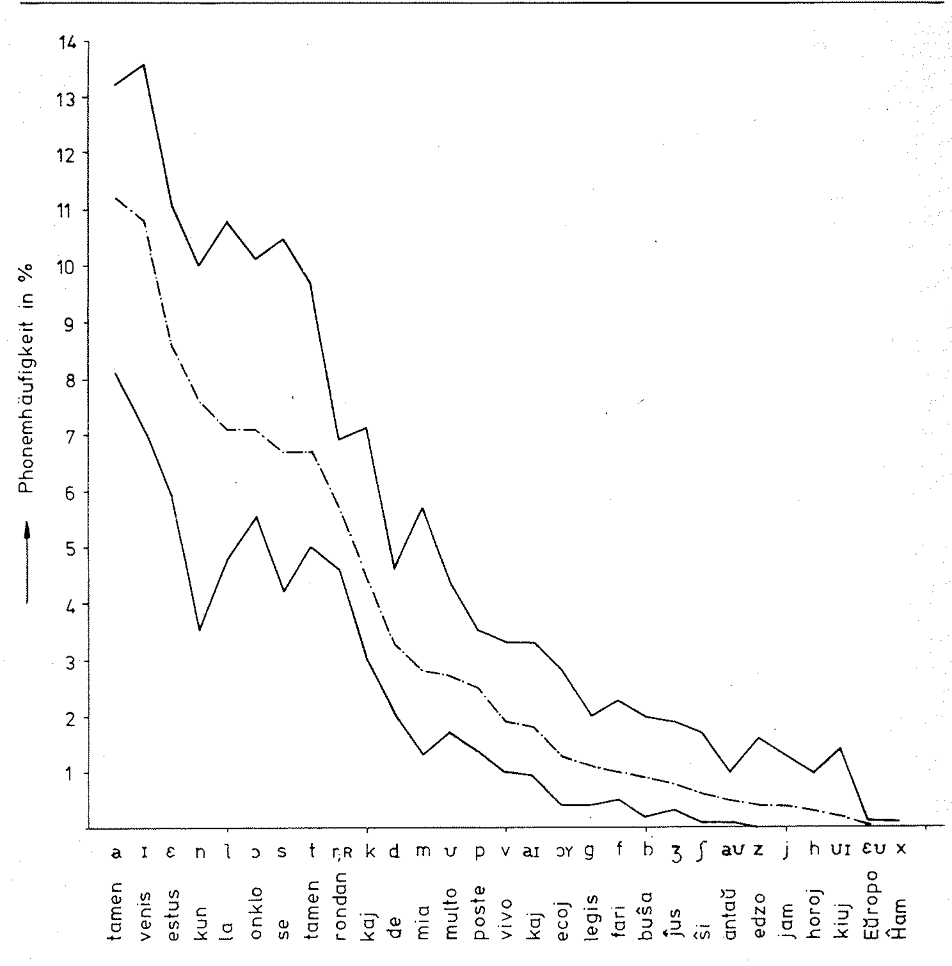


Bild 2: Phonemhäufigkeitsverteilung des Esperanto: Mittelwerte aus 50 000 Phonemen (10 verschiedene Textbeispiele), Extremalwerte aus Teiltextzählungen (1000 Phoneme): Zusammenfassung der Ergebnisse aus Bild 1.

wert ein, der für das betreffende Phonem und für die betreffende Textprobe aus der 5000-Phoneme-Untersuchung hervorgegangen ist. Die einzelnen Ordinatenstrecken geben damit Mittelwerte und den weitesten festgestellten Streubereich bei noch ausreichend langen Textproben für jeden untersuchten Text und für jedes Phonem des Esperanto wieder. Für jedes Phonem sind, textunabhängig, weiterhin die äußersten überhaupt ermittelten Streuwerte durch horizontale Begrenzungen angegeben. Zwischen diesen Begrenzungen sind also jeweils sämtliche aus den zehn untersuchten Textbeispielen erhaltenen 50 Teiltextergebnisse (aus je 1000 Phonemen) eingeschlossen.

Was ist aus den dargestellten Ergebnissen zu folgern? Zwischen den zehn untersuchten Texten sind weder bei den zehn verschiedenen Autoren bzw. den zehn verschiedenen Übersetzern, noch bei den sieben verschiedenen Ursprungssprachen wesentliche allgemeine und eindeutige Unterschiede festzustellen. Am meisten weicht noch der Text 10 (Bibel) von den Ergebnissen der anderen Texte ab. Bei diesem Textbeispiel liegen bei 7 der 29 verschiedenen Phonemen die aus der Untersuchung der gesamten Textprobe von 5000 Phonemen errechneten Mittelwerte der Phonemhäufigkeit deutlich über den übrigen Mittelwerten (aus anderen Textbeispielen) desselben Phonems. In immerhin 11 Fällen bilden die oberen Extremalwerte dieses Textes 10 die höchsten überhaupt bei den betreffenden Phonemen ermittelten Extremalwerte. Bei der Auswahl dieses Zähltextes wurden die Kapitel mit überproportional starken Häufungen bestimmter Textwörter in der Schöpfungsgeschichte (und, Gott, Himmel, Erde) bewußt nicht herangezogen. Trotzdem kann vermerkt werden, daß in der ausgezählten Textprobe (Kapitel 6–8 des 1. Buch Mose) diese Wörter dennoch betont häufiger als die übrigen auftreten. Eine Nachprüfung ergab jedoch, daß von den in diesen Wörtern enthaltenen Phonemen /a/, /I/, /ɛ/, /n/, /l/, /ɔ/, /t/, /R/, /k/, /d/, /ʃ/ (und = /kaI/, Gott = /diɒ/, Himmel = /tʃIɛlɔn/, Erde = /tɛRɔn/) nur die Phoneme /k/ und /ʃ/ zu den in den Darstellungen des Bildes 1 auffälligen Phonemen gehören. Eine abgesicherte Erklärung für die Ursache der beim Zähltext 10 festgestellten geringfügigen Abweichungen gegenüber den Ergebnissen aus den anderen Zähltexten ist also nicht zu geben.

Eine gewisse Aussagekraft besitzt die für jedes einzelne Phonem ermittelte Streubreite der Phonemhäufigkeit als Ergebnis der Untersuchung von Teiltexten aus je 1000 Phonemen Länge für jede der zehn ausgewählten Textproben. Wie Bild 1 zeigt, sind die an den zehn Textbeispielen ermittelten maximalen Streubreiten der Phonemhäufigkeit innerhalb der Ergebnisse für je ein bestimmtes Phonem durchaus nicht einheitlich groß. Bestimmte Befunde (sehr große oder sehr kleine Streubreite) sind jedoch nicht an eine bestimmte Textprobe gebunden, sondern sind statistisch verteilt. Die – willkürlich – gewählte Länge der Teiltexte von je 1000 laufenden Phonemen ist für eine Untersuchung der Textabhängigkeit offensichtlich noch zu klein. Die aus allen 50 Teiltexten je Phonem gewonnene maximale Streubreite der Phonemhäufigkeit läßt jedoch bereits allgemeine Schlüsse zu. Analog zu anderen Zählergebnissen (Meier, 1967; Chavasse, 1948) ist die maximale Streubreite von der absoluten Größe der mittleren Phonemhäufigkeit abhängig. Verglichen mit den von Meier für die deutsche Sprache ebenfalls mit Teiltexten aus je 1000 Phonemen an insgesamt 100 000 ausgezählten Phonemen ermittelten Streubreite sind die für die Phonemhäufigkeit des Esperanto gefundenen Streubreiten etwa halb so groß wie die für die deutsche Sprache. Der Grund hierfür ist offensichtlich das gegenüber dem Deutschen geringere Phoneminventar des Esperanto. Die reichere Differenzierungsmöglichkeit der Phonologie des Deutschen verursacht offenbar bei gleicher Teiltextlänge der untersuchten Sprachproben beider Sprachen eine größere Streuung der Phonemhäufigkeitsteilergebnisse als im Esperanto.

Eine vereinfachende und zusammenfassende Darstellung der Einzelergebnisse aus Bild 1 liefert Bild 2. Es ist hier wieder die Phonemhäufigkeit als Funktion der 29 verschiedenen Phoneme des Esperanto aufgetragen. Auf der Grundlage des gesamten untersuchten Sprachmaterials von 50 000 Phonemen ist ohne Differenzierung nach einzelnen Texten die mittlere Phonemhäufigkeitsverteilung des Esperanto dargestellt. Außerdem sind die höchsten und niedrigsten Werte der Phonemhäufigkeit angegeben, die für die verschiedenen Phoneme bei der Untersuchung aller zehn Texte bei einer Teiltextlänge von je 1000 laufenden Phonemen gefunden wurden. Die Phoneme auf

der Abszissenachse sind wieder, genau wie in Bild 1, nach fallender mittlerer Häufigkeit im Esperanto angeordnet. Unterhalb der Phonembezeichnungen der Abszissenachse sind für die betreffenden Phoneme entsprechende Beispielwörter angegeben.

Die Programme zur Ermittlung der Zählergebnisse und die entsprechenden Programmläufe zur Phonemzählung mit einer Datenverarbeitungsanlage sind im Rechenzentrum der Forschungsgruppe 'Mathematik für Nachrichtentechnik' des Forschungsinstitutes der Deutschen Bundespost, Darmstadt, entwickelt bzw. durchgeführt worden. Ich möchte hier deren Leiter, Herrn Diplom-Ingenieur E. Theissen, für die von ihm geleistete Arbeit und die gute Zusammenarbeit herzlich danken.

Schrifttum

CCIT (Comité Consultatif International des Communications Téléphoniques à Grande Distance): Recueil de listes de logatomes Esperanto pour mesures de netteté. Paris: o.J.; abgedruckt in: CCIT, 8. Vollversammlung, Paris (Rotbuch), 69-89 (1931).
Cervantes Saavedra, M. de: La malprudenta scivolulo. Valencia: L. Hernandez 1955.
Chavasse, P.: Essai sur la phonétique statistique de la langue française et son application a l'étude de l'intelligibilité d'une conversation. Annales des Télécommun. 3, 5-23 (1948).
Chesterton, G.K.: La naiveco de pastro Brown. Heronsgate, Rickmansworth (England): The Esperanto Publishing Co., Ltd. 1937.
Denes, P.B.: On the statistics of spoken English. J. Acoust. Soc. Amer. 35, 892-904 (1963).
Dewey, G.: Relative frequency of English speech sounds. Harvard studies in education, Vol. IV, 2. Aufl. London: Cambridge, Harvard University Press 1950.
French, N.R.; Carter, C.W.; Koenig, W.: The words and sounds of telephone conversations. Bell Syst. Techn. J. 9, 290-324 (1930).
Hamsun, K.: Victoria. Oslo: Norvega Esperantista Ligo 1938.
Hölscher, E.: Esperanto, die internationale Sprache. Nürnberg: Deutscher Esperanto-Bund e.V. 1965.
Immermann, K.L.: La karnavalo kaj la somnambulino. Heronsgate, Rickmansworth (England): The Esperanto Publishing Co., Ltd. 1952.
Janot-Giorgetti, M.T.; Lamotte, M.: Système de reconnaissance automatique des fautes de prononciation. grkg/Humankybernetik 23, 81-90 (1982).
La Rochefoucauld, F. Duc de: Pripensoj aŭ sentencoj kaj primoraj maksimoj. Paris: Librairie Felix Alcan 1935.
La sankta biblio. London: British Bible Society 1954.
Malécot, A.: Frequency of occurrence of French phonemes and consonant clusters. Phonetica 29, 158-170 (1974).
Maupassant, G. de: La normandaj rakontoj. Stockholm: Eldona societo Esperanto 1953.
Meier, H.: Deutsche Sprachstatistik. 2. Aufl. Hildesheim: Georg Olms Verlagsbuchhandlung 1967.
Sotscheck, J.: Ein Reimtest für Verständlichkeitsmessungen mit deutscher Sprache als ein verbessertes Verfahren zur Bestimmung der Sprachübertragungsgüte. Der Fernmeldeingenieur 36, 1-84, Heft 4/5 (1982).
Strindberg, A.: Insulo de feliĉuloj. Leipzig: Ferd. Hirt 1926.
Voltaire, F.M.: Tri verkoj de Volter. Paris: Sennacieca Asocio Tutmonda 1956.
Wagner, H.: Die Weltsprache Esperanto. Stuttgart: Verlag Freizeit und Wandern 1964.
Zweig, S.: La okuloj de la eterna frato. Köln: Heroldo de Esperanto 1932.

Eingegangen am 1. Oktober 1983

Anschrift des Verfassers:
Dr.-Ing. Jochem Sotscheck, Forschungsinstitut der Deutschen Bundespost, Forschungsgruppe Akustik, Ringbahnstr. 130, D-1000 Berlin 42

Counting the phoneme frequency distribution of Esperanto (Summary)

For evaluating the linguistic properties of Esperanto logatoms (test words containing special phoneme combinations generally used for speech intelligibility measurements in telecommunications), it was deemed helpful to compare the properties of these test words with those of various natural languages. It was also felt that the artificial language Esperanto should be included in this comparison. The most important features of interest in this study are the constitution of the set of distinctive speech sounds (phonemes) and their frequencies of occurrence. This paper deals with the findings of an appropriate linguistic analysis of the Esperanto language.

In order to investigate the phoneme frequency distribution of Esperanto, samples consisting of 5 000 phonemes were taken from ten selected texts of the Esperanto literature. The samples were translations from seven languages and had originally been written in very different periods.

The counted results were separately evaluated both for the full length (5 000 phonemes) of all samples and for passages consisting of 1 000 phonemes. These investigations gave no evidence of major differences in the frequency of phoneme occurrence which can be attributed to the individual original texts.

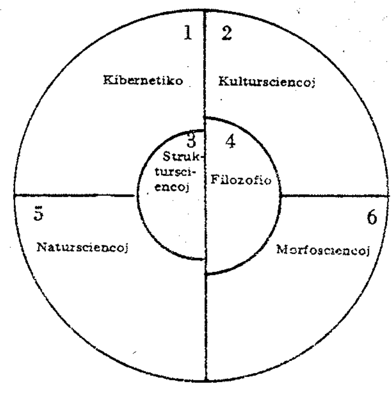
The partial results were combined to obtain a general mean phoneme frequency distribution of Esperanto The spread of the frequency countings of the different phonemes, which has been derived from the study of texts each consisting of 1 000 phonemes, was shown to have only about half the width of the frequency deviation found in German texts. The decrease in the width of the individual phoneme frequency countings is assumed to be attributable to the fact that the phoneme set of Esperanto is considerably smaller than that of the German language.

Noto pri proponita sciencoklasigo por strukturigi sciencan akademion

de Helmar FRANK, Paderborn (D)

Kutime la strukturigo de universitatoj kaj sciencaj akademioj sekvas (se entute!) sciencoklasigon, kiu preteratentas la principan memstarecon de la kibernetiko kaj la daŭran aktualecon de fenomenologiaj realaĵsciencoj (nomindaj „morfosciencoj“). Krome ili ofte emas kunigi la matematikon kun la natursciencoj kaj la filozofion kun la kulturscienсoj. La statupropono (1983/1683 pfR - 09 - 26) por la Akademio Internacia de la Sciencoj San Marino aplikas scienсteorie pli kontentigan klasigon por strukturigi la Akademion en ses sekciojn je po 3-4 fakaroj (vd. grkg/Humankybernetik 3/83, p.151-152). Ĝi estas skizita per la sekva kvara apendico de tiu nia propono.-

La angla kaj franca vorto „science" inkluzivas nur la sciencojn kiuj enkondukas kalkulojn kaj celas prognozojn (t.n. „nomotetaj" aŭ „parametraj" aŭ prospektivaj" sciencoj de la sekcioi 1, 3 kaj 5). La internacilingva termino „scienco" (same kiel la germana termino „Wissenschaft") inkluzivas krome ankaŭ la t. n. „ideografajn" aŭ „neparametrajn" sciencojn (sekcioj 2, 4 kaj 6), kiuj apenaŭ uzas matematikan modeligon, sed strebas al kompreno de la fenomenoj ankaŭ surbaze de regantaj valoroj. - AIS flegas ĝis la limoj de esperebla intersubjekteco ambaŭ alojn de la tuta sciencaro. La aldonita bildo reprezentas la strukturigon de AIS en sekciojn kaj iliajn najbarecojn. La unnopaj sekcioj estas provizore subdividitaj en fakarojn laŭ la sekva klasigo.



1. *Kibernetiko (pritraktanta mezurante kaj modeligante informon)*

1.1 *Antropokibernetiko (psikokibernetiko, lingvokibernetiko, kibernetika estetiko, soci- kaj ekonomikibernetiko, klerigkibernetiko, jurkibernetiko).*

1.2 *Ĝenerala kibernetiko (teorio de informacio kaj enkodigo, teorio de logikaj ŝaltretoj kaj abstraktaj aŭtomatoj, ĝenerala reguladteorio, teorio de adaptiĝaj kaj celorientitaj sistemoj, informadiko).*

1.3 *Maŝinkibernetiko (teknologioj de telekomunikado, de informstorado, de reguligo, de komputiloj, kaj de la aŭtomacio).*

1.4 *Biokibernetiko (inkluzivante la kibernetikan medicinon).*

2. *Kulturscienсoj (pritraktantaj informecajn fenomenojn apenaŭ aplikante mezuradon kaj matematikan modeligon)*

2.1 *Bazaj nekibernetikaj antroposcienсoj (neparametraj psikologio, etnologio, lingviko, kaj sociologio).*

2.2 *Sociscienсoj nekibernetikaj (administracischienco, ekonomiko, jurscienco, politologio, neparametra klerigscienco, nekibernetika komunikad- kaj dokumentadscienco).*

2.3 *Artscienсoj neparametraj (muzikologio kun muzikhistoriko, estetiko kaj historiko de la porvidaj artoj, beletriko, estetiko kaj historiko de la scenejaj artoj).*

2.4 *Integraj kulturscienсoj (kompara religiscienco kun religihistoriko, socigeografio, ĝenerala historiko).*

3. *Strukturscienсoj (analizantaj kaj formale modeligantaj la mekanismojn de la aprioraj ekkonoj)*

3.1 *Pensadstrukturscienсoj (simbola logiko, teorio de algoritmoj, ĝenerala aksiomatiko, aritmetiko kaj nombroteorio, probablokalkulo kaj matematika statistiko).*

3.2 *Imagadstrukturscienсoj (geometrio, kinematiko, ĝenerala stistemiko).*

3.3 *Formalaj matematikaj scienсoj (algebro, teorioj de serioj, funkcioj, ekvacioj, distribucioj ktp., kalkuloj de vektoroj, matricoj, diferencialoj, integraloj ktp.).*

4. *Filozofio (celante unuflanke plej bazan, aliflanke kunigante-kompletigan komprenon de la esto, iĝo kaj ekkono)*

4.1 *Filozofiaj bazoscienсoj (filozofia logiko, epistemologio)*

4.2 *Filozofiaj valorteorioj (filozofia estetiko, moralfilozofio, jurfilozofio, ŝtatfilozofio).*

4.3 *Model- kaj sciencteorioj (ĝenerala modelteorio, ĝenerala sciencteorio, filozofioj de la unuopaj scienсoj).*

4.4 *Idehistoriko (historiko de la filozofio, historikoj de la unuopaj scienсoj).*

5. *Naturscienсoj (pritraktantaj mezurante kaj modeligante materion kaj energion)*

5.1 *Substancikoj, t.e. ĝeneralaj scienсoj pri la koncernaj, loke kaj tempe invariantaj leĝoj (fiziko kun siaj teknologioj: konstrustatiko, maŝinteknologio, elektroteknologio ktp.; kemio kun siaj teknologioj).*

5.2 *Mondikoj, t.e. geo-kaj astroscienсoj (astronomio, geodezio, naturgeografio, geologio).*

5.3 *Bioscienсoj (ĝenerala biologio, botaniko, zoologio, medicino, agronomio, nutradscienco, ekologio).*

6. *Morfoscienсoj (celantaj pli intuicie ol matematike ĉerpi informon el la observita materieca-energieca mondo kaj formi ĝin konforme al ideoj)*

6.1 *Scienсoj pri aŭdvida dokumentado (sciencaj fotado, filmado, sonsurbendigo ktp.; museologio).*

6.2 *Ilarmorfologio (ergonomiko, formado de industriaj produktoj).*

6.3 *Mediplanado (arkitekturo, ĝardenformado, urbo- kaj regioplanado, trafikplanado, turismiko).*

grkg/Humankybernetik
Band 24 · Heft 4 (1983)
Gunter Narr Verlag Tübingen

Probleme der semantischen Analyse bei der automatischen Faktenrecherche

von Georg F. MEIER, Berlin (DDR)

Sieht man von rein graphischen Informationen (Diagrammen, Tabellen u.ä.) ab, so gilt, daß jede Information mittels sprachlicher Einheiten erfolgt. Im allgemeinen handelt es sich um mehr oder minder umfangreiche textuelle Äußerungen, die gegebenenfalls in kleinere inhaltstragende Einheiten zerlegt werden können: Text – Absatz – komplexer Satz – einfacher Satz bzw. Teilsatz (Gliedsatz) – Satzglied – Syntagma – Wort – (selten) Wortteil (d.i. Monem, Radikal, Stamm, Wurzel usw.). Die Zergliederung (parsing) ist in vielen Sprachen formal durch Absatz, Interpunktion, Großschreibung (im Deutschen), Satzgliedpartikel (z.B. *ga* und *wa* im Japanischen), Artikel (in einigen europäischen Sprachen, im Arabischen, Ungarischen) vorbereitet. Auch wenn Interpunktsregeln keineswegs "logisch" sind, so sind sie doch festgelegt und programmierbar. Selbst die Wörter (Lexeme) und Moneme (Morpheme) können mehr oder minder vollständig im Permanentspeicher (Lexikon) eines Rechners erfaßt und damit abrufbar gemacht werden. Die Reihenfolge der Lexeme innerhalb einer Texteinheit kann abgezählt werden und mit eingespeicherten Distributionsregeln verglichen werden. Es kann somit eine formale Syntax erkannt werden, die ihrerseits je nach Sprachsystem bestimmte grobe Aussagen über die inhaltliche Struktur zuläßt. All diese Operationen sind heute für einige Sprachen – sieht man von einigen Problemen ab – prinzipiell gelöst, sofern die entsprechende Kleinarbeit (vollständiges Lexikon, vollständiges grammatisches Regelsystem, evtl. Transliteration usw.) vorliegt. Trotzdem ist weder eine automatische Übersetzung von hoher Qualität noch eine Speicherung von Sachverhalten bisher gelungen, obwohl besonders für die Übersetzung eine Reihe von Zuordnungsoperationen, teilweise ganzer Phrasen, gefunden wurden, so daß eine eigentliche Inhaltserkennung ausgespart werden kann. Bei einer Mehrfach-Sprachübersetzung, d.h. bei sogenannten multilateralen Verfahren ist eine "Zwischensprache" (Metasprache, prädikatenlogische Sprache, Interlingua, Informationslogisches System) erforderlich, da dieses Verfahren im Prinzip reversibel funktionieren muß. Begriffskataloge, wie numerische und nicht-numerische Klassifikationssysteme, Facettensysteme, Thesauren unterschiedlicher Präzision, Informationsrecherchesprachen, wie SYNTOL oder der Univerzal'nyj semantičeskij kod V. Martynovs, Noematika u.a. sind sicher eine wichtige Hilfe für die Umkodierung sprachlicher Lexeme auf ihre Sememe und Bedeutungselemente, aber sie stellen noch keine Basis für eine Sachverhaltsdarstellung her. Versuche auf Grund der Häufigkeit der jeweiligen Umgebung gewisse Kerninhalte zu erfassen, setzen entsprechende statistische Untersuchungen und Distributionsklas-

sen voraus, die nur auf der Basis bereits bekannter Informationen geschaffen wurden, aber eine Neuinformation unwahrscheinlich erscheinen lassen. Frage-Antwort-Systeme, die im Bereich komplexen Wissens wirksam werden sollen, d.h. über den Bereich beschränkter Auskunftstexte hinausgehen, müssen exakte Inhaltsspeicherungen besitzen, die nicht mit den realen Sätzen natürlicher Sprachen identisch sind. Für Dokumentenrecherchen mögen Deskriptoren bzw. Schlüsselwortverfahren genügen, obwohl auch hier erst durch entsprechende Umgebungselemente die Relevanz des Nachweises erfaßt werden kann. Gedankliche Zusammenhänge, Werturteile, Entscheidungsgrundlagen, Sachverhaltsabfragen und Wissenskombinationen sind nur möglich, wenn das Erschließen des Inhalts eines in einer natürlichen Sprache vorliegenden Textes möglich sein wird. Spätestens um 1970 wurde international die Unlösbarkeit von Informationsspeicherung auf asemantischer Basis erkannt Damit setzte sich die von verschiedenen Seiten schon in den 60er Jahren geforderte Konzentration auf die semantische Analyse als wichtige Voraussetzung für diesen Zweig der linguistischen Datenverarbeitung immer mehr durch (vgl. Meier, 1966). Zwanzig Jahre später kann man erst von Anfangsergebnissen sprechen, von einer Vielzahl von Methoden, teils nur auf Minimaldistinktion bedacht (Antonymverfahren), teils mehr psychologisch fundiert (skalare Methoden), teils nur innersprachlich angelegt (sog. "Sem"-Analysen), teils intersprachlich aufgebaut (noematische Methoden), teils mehr auf Satzanalysen unter Verzicht auf Sememdefinitionen, teils nur auf Methoden der Formalisierung ausgerichtet. Die meisten Verfahren – soweit sie nicht nur zur Stützung einer Grammatik- oder Semantik-Hypothese dienen sollten – weisen einen hinreichenden Grad adäquater Methodologie und Selbständigkeit auf. Es ist weder im Rahmen dieses Aufsatzes noch beim gegenwärtigen Entwicklungsstand überhaupt ratsam, eine Zusammenschau aller semantischen Verfahren vorzunehmen. Es geht uns hier vielmehr darum, Probleme zu zeigen und auf Lösungsmöglichkeiten aus dem eigenen Erfahrungsbereich zu verweisen.

Wir gehen von zwei allseitig bekannten Fakten aus: 1) alle Menschen können die in ihrer Muttersprache oder einer relativ gut erlernten Sprache geäußerten Gedanken (Sachverhalte, Fragen usw.) verstehen, sind also zur semantischen Analyse der Äußerungen fähig; 2) alle Lexikographen, ebenso die Verfasser einsprachiger wie zwei- oder mehrsprachiger Wörterbücher, mußten und konnten intuitiv die Bedeutung(en) der erfaßten Wörter analysieren. In beiden Fällen wirkten und wirken die kontextuellen Elemente mit, sei es um die jeweilige aktuelle Bedeutung zu erkennen oder sei es aus den Kontexten eines Wortes seine potentiellen Bedeutungen zu notieren. Der sehr komplizierte Mechanismus unserer Gehirnarbeit ist trotz verschiedener blackbox-Experimente noch kaum erschlossen, so daß ein bionisches Modell Hypothese bleibt. Da die Sprache als Ergebnis vieler teils sehr weit zurückliegender Perioden ihres historischen Werdeganges aus allen Perioden Residuen und Neuerungen in sich birgt, ist sie mit Mitteln der Logik nicht beschreibbar. Je mehr Sprachen zu einem lexiko-semantischen Vergleich herangezogen werden, desto deutlicher läßt sich der große Unterschied in der jeweiligen Bedeutungszuordnung erkennen. Dasselbe gilt für die formale und semantische Seite der Grammatik. Es gibt nahezu keine Universalien in der Syntax, Morphologie, bei Wortarten oder grammatischen Kategorien. Die semantische Analyse von Sätzen fällt von Sprache zu Sprache ganz anders aus. So kann man nicht von einer scheinbar logischen oder ontologischen Kategorie auf die einer anderen Sprache schließen. Ebenso ist es unmöglich ein Wort in der Sprache L_1 eineindeutig auf ein

Wort in der Sprache L_2 abzubilden. Das hat nicht nur zur Folge, daß vor einer Abbildung eines Lexems auf ein anderes die Polysemie ausgelöscht wird, d.h. daß durch Auffinden der richtigen aktuellen Bedeutungen aus der Zahl der möglichen Bedeutungen eines Wortes oder einer Wortgruppe (Syntagma) das jeweilige abbildbare Semem festgelegt werden muß, sondern daß durch exakte Definition der jeweiligen Bedeutung (Semem) auch die Bedeutungsbreite zum Vergleich bereitgestellt werden muß.

Wenden wir uns zunächst dem letztgenannten Problem anhand einiger einfacher Bedeutungsfelder (Bedeutungsklassen) zu: Für genealogische oder historische Informationen spielen die Verwandtschaftsbeziehungen eine nicht unwichtige Rolle. Auch für juristische Fragen (Erbschaftsfragen usw.) kann eine genaue Bestimmung wichtig sein. Relativ universal sind – soweit patriarchalische Verhältnisse bestehen – die Sememe 'Vater' und 'Mutter' (rein genealogisch) unter Ausschluß von Stiefverwandtschaften und Adoptiv-Verwandtschaften. Aber es gibt Sprachen, die Unterschiede für den *eigenen* Vater und für den *fremden* Vater machen oder – wie im Tibetischen – mit *pha* den 'gewöhnlichen Vater' und mit *jab* den 'Vater einer Respektsperson' bezeichnen. Schon bei den Kindern ist die Differenzierung unterschiedlich: so ist im Georgischen *sv̌ili* 'Kind' und 'Sohn', während die 'Tochter' als *kalisv̌ili* bezeichnet wird (Frau-Tochter), ähnlich span. *hijo* (Sohn)/*hija* (Tochter), aber *hijos* (Kinder und Söhne). Noch komplizierter wird die Abbildung bei Geschwistern. So sind im Ungarischen älterer Bruder (*bátya*) und jüngerer Bruder (*öcz*) und früher auch bei den Schwestern *néne* und *hug* unterschieden worden, bei den Mongolen *aha* (älterer Bruder) und *düü* (jüngerer Bruder), ähnlich im Vietnamesischen u.a. Sprachen. Im Baskischen und Svanischen (im Kaukasus) muß zwischen dem Bruder des Bruders (*anai*) und dem Bruder der Schwester (*neba*) unterschieden werden, d.h. eine Schwester nennt ihren Bruder anders als der Bruder seinen Bruder nennt. Desgleichen nennt die Schwester ihre Schwester *ahizpa*, während der Bruder sie *arreba* nennt. Svanisch analog nennt die Schwester ihre Schwester *udil* und ihren Bruder *ǧemil*, der Bruder seine Schwester *dač̕wir* und seinen Bruder *muxwbe*. Während Deutsche, Engländer, Franzosen, Italiener nur zwischen 'Onkel' und 'Tante' bzw. *uncle/aunt* bzw. *oncle/tante* bzw. *zio/zia* unterscheiden, differenziert das Schwedische zwischen *farbror* (Vaterbruder) und *morbror* (Mutterbruder) für den 'Onkel' und analog *faster* bzw. *moster* für die 'Tante'. Diese Unterscheidung treffen auch die Araber *ʿamm* (Vaterbruder)/*ʿamma* (Vaterschwester) bzw. *xāl* (Mutterbruder)/*xāla* (Mutterschwester) oder – sicher teils durch offenbaren Einfluß des Arabischen, teils original – im Aserbajdschanischen *ämi* (Vaterbruder)/*bibi* (Vaterschwester) bzw. *dady* (Mutterbruder)/*xala* (Mutterschwester). Das Serbokroatische unterscheidet auch zwischen Vaterbruder (*stric*) und Mutterbruder (*ujak*), aber auch dem Mann der Vaterschwester oder Mutterschwester (*tetak*). Im älteren Slowakischen hat man u.a für die Frau des Mutterbruders *ujčina* und des Vaterbruders *stryná*. Andere Sprachen unterscheiden zwischen den verschiedenen verschwägerten Verwandten, so bulgarisch *šurej* (Schwager des Ehemanns), *dever* (Schwager der Ehefrau), *zălva* (Schwägerin der Ehefrau), *baldăza* (Schwägerin des Mannes). Im Vietnamesischen wird u.a. noch zwischen dem älteren Brudes des Mannes (*anh chồng*), dem älteren Bruder der Ehefrau (*anh vợ*), dem Mann der älteren Schwester (*anh rễ*), zwischen jüngerer und älterer Cousine, zwischen jüngstem Bruder und jüngerem Bruder usw. unterschieden. Die Beispiele mögen genügen, um zu zeigen, daß familienrechtliche oder kultusbezogene Gründe die eine oder andere Differenzierung erforderlich mach-

ten. Bei einer einspeicherbaren Verwandtschaftseinheit genügen also nicht Sexus, Generation, Blutsverwandtschaft, Sonst entstehen – wie das Beispiel 'Schwager' oder 'Onkel' zeigt – immer Schwierigkeiten, wenn man von der weniger differenzierenden in die enger differenzierende Sprache übersetzt, da hier Informationsmangel vorliegt.

Handelt es sich hier um unverkennbare gesellschaftliche Gründe, so ist dies z.B. bei der Bezeichnung von Farben sicher nicht der Fall. Lassen sich Verwandtschaftsgrade relativ leicht als prädikatenlogische Kalküle formalisieren, z.B. sohn = R_{pat} (a,b) + b ϵ mask, so macht der ideelle Charakter der Farbempfindung besondere Schwierigkeiten, selbst wenn man bei Elementar-Farben die physikalischen Daten nennen könnte (also Reflexionskoeffizient, Wellenlänge und spektrale Farbdichte). Selbst wenn wir nur 10 000 Farben unterscheiden könnten – es gibt viel höhere Schätzungen (Kries und Judd setzten mindestens 300 000 an), so stehen dem höchstens 200 bis 250 Farbbezeichnungen, viele mit Vergleichselementen, wie *fuchsrot, smaragdgrün* usw. gegenüber. Schlimmer aber ist es, wenn – wie in der Munda-Sprache – nur drei Farbwörter vorhanden sind oder die Farbbezeichnungen von Sprache zu Sprache sich nicht decken, wie z.B. bei chinesisch *qing*, das die Skala grasgrün, teegrün, olivgrün, bleifarben, grau, blau bis schwärzlich umfaßt oder wenn dem tschuwaschischen *kavak* (aus altturksprachlich *kök*) im Russischen *seryj, sizyj, goluboj, sinij, sedoj, sivyj* und *serovatyj* entsprechen, d.h. daß die Farbe *kavak* die Breite von tiefgrau, bleigrau, hellblau, tiefblau, schimmelgrau, silbergrau, grauweiß bestreitet. Der Computer kann hier nicht strenge Grenzen ziehen, zumal – wie im Chinesischen und anderen Sprachen Synonyme existieren, die jedoch nur in bestimmten Syntagmen verwendet werden können, so wird bei 'grün' für Wald das Wort *lü* verwendet, das auch für Bohnen und schwarze Haare benutzt werden muß. Eine Basis kann der *Dictionary of Color* (erstmalig 1930 mit 4000 Farbenbezeichnungen erschienen) bilden, doch müssen für die Synthese z.B. ins Chinesische oder Japanische die entsprechenden Umgebungen programmiert sein.

Besondere Unterschiede in der Bedeutungsbreite bzw. im syntagmatischen Anwendungsbereich weisen auch viele Verben bzw. Prozeßbezeichnungswörter auf. So hat das deutsche Verbum *abnehmen* rund 40 verschiedene Übersetzungsentsprechungen im Italienischen, während die entsprechenden italienischen Verben keineswegs im Deutschen nur durch *abnehmen* zu übersetzen sind (vgl. G.F. Meier, 1978, pp.143-150). Das trifft nun für nahezu alle anderen Sprachen in ähnlicher Weise zu, z.B. um 35 verschiedene Übersetzungsäquivalente im Russischen, wobei z.B. das deutsche *abnehmen* von Hut, Mantel, Ring im Russischen durch *snjat'* wiedergegeben wird, das seinerseits auch für das *ausziehen* bei Schuhen oder anderen Kleidungsstücken verwendet wird oder für photographieren, wofür wir höchstens *aufnehmen* (nicht *abnehmen*) sagen können. Um 35 verschiedene deutsche Verben müssen für *snjat'* verwendet werden, so *abbrechen* (von Brücken), *entfernen, abputzen, aufnehmen, belegen, mieten, abheben, verhören, streichen, zurücknehmen, absetzen* u.a. In der gleichen Größenordnung verhält sich Deutsch zu anderen slavischen Sprachen. Für die Analyse muß also schon ein Abfrage-Algorithmus vorbereitet sein, der zunächst nach intransitiver Verwendung (Fehlen des Objekts) und dann nach der Bedeutung des Subjekts, also z.B. *Mond, Mensch, Fieber, Kraft, Vorräte . . .* fragt und im Falle der Besetzung der Objektsstelle nach der Bedeutung des Objekts fragt, also *Eid, Parade, Koffer, Verantwortung, Last, Wäsche, Telephonhörer* usw. Dabei muß der Algorithmus soviele Fragen stellen, wie es sich aus der Summe aller beteiligten Sprachen ergibt.

Ein weiteres Problem stellen die *festen Syntagmen* dar oder zunächst noch die Feststellung welcher Teil eines Satzes ein im Lexikon eingetragener Bedeutungsträger ist. Dabei müssen wir zwischen Komposita und Komplexwörtern unterscheiden. Im Deutschen oder Ungarischen werden Nominalkomposita zusammengeschrieben, so daß sie als ein Lexem leicht erkennbar sind. Aber in anderen europäischen und asiatischen Sprachen muß erst das Formativ, d.h. der eigentliche Bedeutungsträger erkannt werden. Hier einige Beispiele:

deutsch *Bildkristallgleichrichter*
enlglisch: *crystal video rectifier* (drei Elemente)
russisch: *kristalličeskij videodetektor* (zwei Elemente)
französisch: *redesseur à cristal vidéo* (vier Elemente)
italienisch: *radrizzatore video a cristallo* (vier Elemente)
spanisch: *rectificador de cristal de video* (fünf Elemente)
japanisch: ji shin ko tei kio ku soo chi (= permanenter Magnetkernspeicher)
(acht Elemente)
vietnames.: máy dao đông có tia đien (= Elektronenstrahl-Oszillograph)
(sechs Elemente)

Die Lösung bestünde in der prinzipiellen Suche nach dem längsten im Lexikon eingetragenen Syntagma, was aber bei morphologiearmen Sprachen zu Verwechslungen zwischen festen und nicht festen Wortverbindungen führen kann. Außerdem muß das rasche Anwachsen des technischen Fachwortschatzes zu ständigen Korrekturen des Lexikons führen.

Das andere Problem stellen Komplexwörter aus verschiedenen Wortarten, besonders sogenannte Streckformen dar. Im Deutschen, aber auch in anderen europäischen Sprachen nehmen diese Formen immer mehr zu, wie etwa *zu Gehör bringen, zur Kenntnis nehmen, zu überlegen geben, Maßnahmen treffen* usw. Aber auch Komplex-Präpositionen wie *in Zusammenhang mit, oberhalb von, unter der Bedingung daß, unter Ausschluß von, zusammen mit,* franz. *au lieu de,* engl. *because of* usw. In manchen Sprachen stellen solche Komplexwörter einen hohen Prozentsatz des Wortschatzes dar, so in den Turksprachen, z.B. aserbajdschanisch *müšahidä etmäk* (Beobachtung machen), *baša düšmäk* (wörtl. in den Kopf fallen – verstehen); in iranischen Sprachen, z.B. persich *tahsil kardan* (Studium machen); in indoarischen Sprachen, z.B. *hindi khetī karnā* (Landarbeit machen), nepali *daga dinu* (Betrug geben = betrügen); japanisch *benkyō suru* (Studium machen), vietnamesisch *vê nhà* (zurückgehen Haus = heimkehren), *di chân* (gehen Fuß = zu Fuß gehen); chinesisch *kaidòng* (in Bewegung setzen), *danxin* (tragen-Gedanken = sich sorgen). Hier greifen Syntax- und Lexikanalyse eng ineinander über, da – besonders beim Deutschen – die Teile im Satz getrennt werden können. Auch diese Komplexwörter müssen im Lexikon mit Verweis auf eventuelle syntaktische Suchregeln eingetragen sein. Das gilt auch für Korrelativa, wie italienisch *tanto . . . quanto,* deutsch *je . . . desto . . .* usw.

Das Hauptproblem stellen aber nach wie vor die lexische und grammatische Polysemie dar. Wir erwähnten oben die Nichtübereinstimmung zwischen den Bedeutungen in verschiedenen Sprachen. Das geht natürlich teilweise auf die Polysemie zurück. Andererseits gibt es auch innerhalb einer Sprache die Polysemie, die der Muttersprachler fast automatisch monosemiert, während der Computer eine immense Arbeit ausfüh-

ren muß, da er zunächst im Lexikon alle möglichen Sememe vorfindet und nun mit Hilfe einiger syntaktischer und speziell lexiko-semantischer Abbfragen die Monosemierungsoperation ausführen muß. Wir wollen an einem relativ einfachen Fall den Monosemierungsprozeß vorführen, wobei einige Vereinfachungen vorgenommen worden sind. Die Prüffragen richten sich nach Satzgliedfragen und Kasus, die zuerst ermittelt werden müssen, um dann die lexischen Besetzungen der Leerstellen ermitteln zu können. Der Satz *Die ersten Spieler nehmen ihren Platz ein* enthält das Verb *einnehmen,* das polysem ist. Um es zu monosemieren muß es eine algorithmisierte Operation durchlaufen, die für alle Sememe von *einnehmen* gültig ist. Vorher muß die syntaktische und grammatisch-semantische Analyse erfolgen. *Spieler* kann Singular oder Plural, Nominativ oder Genitiv-Plural, Dativ-Singular und Akkusativ sein, ebenso ist *ersten* polysem, sogar *die*. Die Satzanalyse erfolgt auf der Basis der Satzgliedanalyse, hier also zunächst Nominalgruppe-Verbum-Nominalgruppe-Präfix. Doch auch dies muß erst festfestellt werden. Im Lexikon sind alle möglichen Wortformen eingetragen, also auch *ersten, ihren, nehmen . . . ein*. Zugleich ist die Wortklasse beigefügt, so daß das Verbum als solches sofort festgelegt ist, ebenso Nominalgruppen.

Wir zeigen hier vereinfacht die syntaktische Entscheidungsoperation, wobei ein Minus in der Spalte die gesamte Spalte annulliert. Treffen zwei positive Spaltenwerte bei nur einer möglichen Lösung zu, so ist ein weiterer Durchgang erforderlich. (Bild 1) Das Ergebnis ist eine klare Satzanalyse, so daß nun nach einem Subjektsnominativ oder Objektsakkusativ (O_4) hinsichtlich der lexischen Besetzung gefragt werden kann.

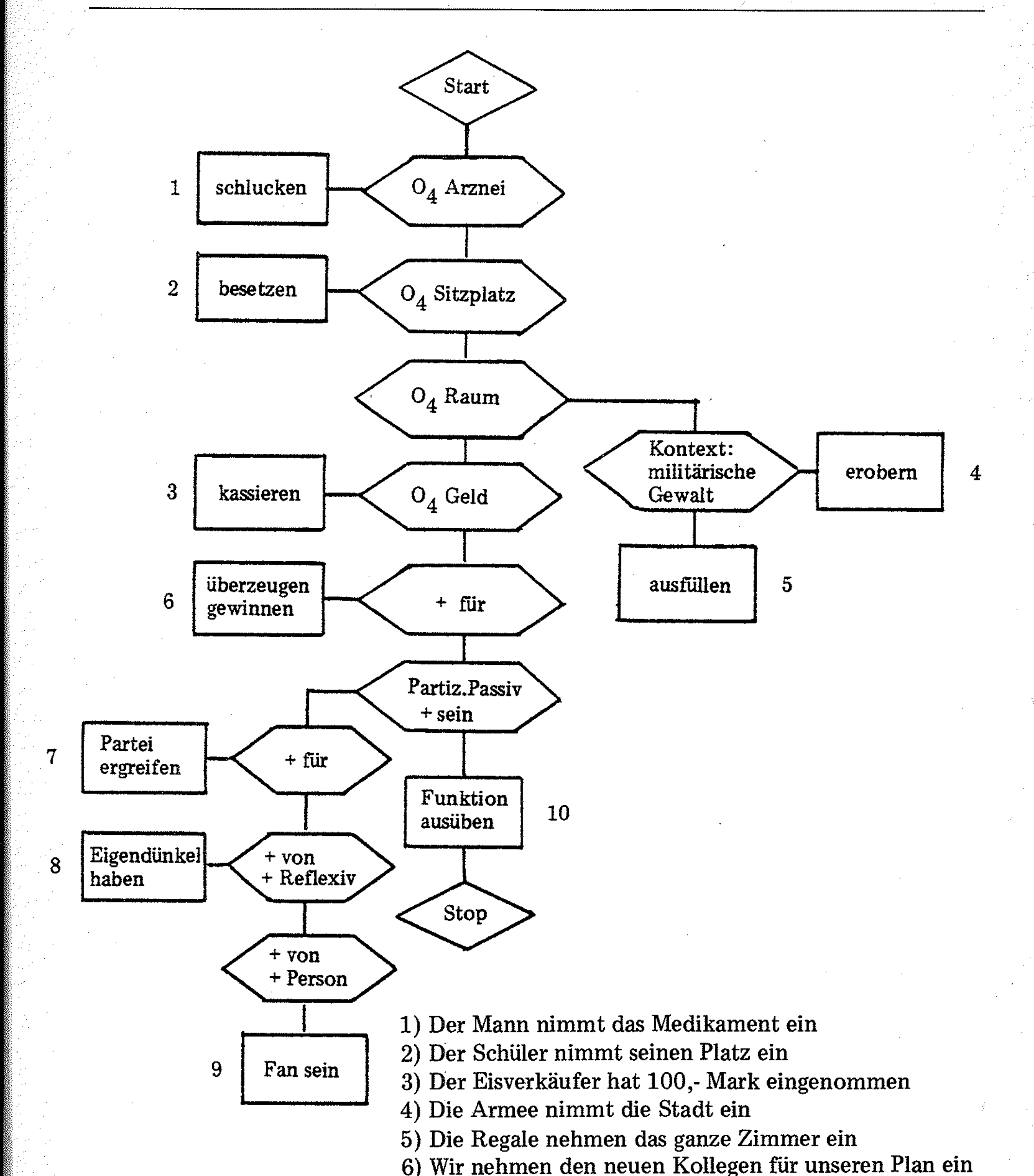
In Bild 2 wird der Algorithmus der Monosemierung gezeigt. Es gibt hier aber Fälle, in denen innerhalb eines Satzes noch keine Monosemierung durchgeführt werden kann. Das betrifft die Frage nach Kontext 'militärische Gewalt'.

Die Entscheidung zwischen dem Semem 4 und 5 ist in dem Satz *Die vier Panzer nehmen den Schulhof ein* ohne weiteren Kontext nicht möglich, da sie im Zuge einer Kampfhandlung oder zur Unterstellung bei einem Manöver in den Schulhof gefahren sind. Im letztgenannten Falle heißt *einnehmen*, daß nur vier Panzer in den Hof passen. Man würde hier gewöhnlich noch das Adjektiv *ganz* einfügen. Wir haben die Kontextfrage deshalb auf 'militärische Gewalt' gegen früher 'Militärbegriff' erweitert, da dieser schon durch 'Panzer' gegeben wäre. Allerdings bedeutet ein solcher Kontextverweis oft eine umfangreiche Liste von Schlüsselwörtern oder Wortgruppen, wie etwa "nach langem Kampf" usw. Glücklicherweise sind solche polysemen Sätze schon theoretisch selten, kommen praktisch noch seltener vor.

Grammatik heißt nicht nur Regelsystem formaler Art, sondern vielfach auch semantische Aussage. Dies gilt für alle grammatischen Kategorien, die nicht nur zur Herstellung von Kongruenzen dienen. Ob es sich um Einzahl oder Mehrzahl von Dingen oder Personen handelt, kann die Maschine nur an sprachlichen Mitteln erkennen. Da die Analyse beim Prädikativ, d.h. meist beim Prädikatsverb beginnt, müßte aus der Verbform die Numeruskategorie und aus ihr die reale Zahl ersichtlich sein. Dies trifft nun aber in vielen Sprachen nicht zu. Es gibt Sprachen ohne Pluralkategorie, wie Chinesisch, Japanisch, Vietnamesisch u.a., wenn man von den Pluralbildungen bei den Personalpronomina absieht; es gibt Sprachen, die am Verb keine Numeruskategorie ausdrücken oder in der 3. Person keinen formalen Unterschied aufweisen (z.B. Litauisch); es gibt Sprachen, die im Prädikat die an sich mögliche Pluralform nur verwenden, wenn das Subjekt Menschen sind, z.B. aserbajdschanisch *mäktäbin häjätindä ušaqlar ojnajyr-*

	I		II			III		IV					
Lexem	Numerus		Genus			Genus - Numerus		Genus - Numerus - Kasus				Verbum	
	Sg	Pl	m	f	n	m Sg	m Pl	N m Pl	G m Pl	D m Pl	A m Pl	tr	itr
die	+	+	+	+	+	−	+	+	−	−	+		
ersten	+	+	+	+	+	+	+	+	+	+	+		
Spieler	+	+	+	−	−	+	+	+	+	−	+		
	+ } → III	+ } → III	+ → III	−	−	−	+ → IV	+ } → Prüfe, ob noch Akkus. im Satz!	−	−	+	+	−
								N m Sg	G m Sg	D m Sg	A m Sg		
ihren	+	+	+	+	+	Ø	Ø	−	−	−	+		
Platz	+	−	+	−	−			+	−	+	+		
	+	−	+								+ → nur Akkus.		

Bild 1



1) Der Mann nimmt das Medikament ein
2) Der Schüler nimmt seinen Platz ein
3) Der Eisverkäufer hat 100,- Mark eingenommen
4) Die Armee nimmt die Stadt ein
5) Die Regale nehmen das ganze Zimmer ein
6) Wir nehmen den neuen Kollegen für unseren Plan ein
7) Sie ist völlig für den Friedenskampf eingenommen
8) Der Vorsitzende ist stark von sich eingenommen
9) Sie ist von diesem Sänger völlig eingenommen
10) Er nimmt die Stelle des Vorsitzenden ein

Bild 2

dyrlar (wörtl. Schule-Gen. auf-Hof-ihren Kind-pl. spiel-imperf.-pl.), aber *bulutlar göj üzünü örtmüšdü* (wörtl. Wolke-pl. Himmel an-ihn bedecken-prät.-sg.). Ähnliches finden wir im Georgischen, in iranischen Sprachen, auch in Bantusprachen. Desgleichen stehen in vielen Sprachen nach den Numeralia entweder Singular oder bestimmte Zählwörter, so türkisch *beş ev* (5 Haus), georgisch *xuti saxli* (5 Haus), baskisch *bortz etxe* (5 Haus), ähnlich ungarisch usw. In anderen Sprachen, wie z.B. den slavischen steht nach 2–4 der Genitiv Singular, ab 5 Genitiv Plural; im Arabischen steht 3–10 mit Substantiv im Genitiv Plural und das Zahlwort im entgegengesetzten Genus, also *thalāthat* (fem.) *fellaḥīn* (m.Gen. Pl.). Im Chinesischen, Burmesischen, Japanischen usw. muß der Plural bei allgemeiner Zahl aus dem Kontext erschlossen werden, während er bei konkreten Zahlen mit einem Zählwort ausgedrückt wird, das je nach Bedeutungsklasse zugeordnet werden muß, so burmesisch -ū für Personen, gău für Tiere, khu für Wörter, šu für Pagoden usw. Bei Ergativsprachen muß das Objekt aus dem Nominativ erschlossen werden, während das Agens im Ergativkasus steht, doch kann dies bei einem Teil der Ergativsprachen noch vom Tempus abhängen. In vielen Sprachen werden Subjekt und beide Objekte in den Verbalkörper inkorporiert, z.B. baskisch *guk saldu geniezazkinakeen* Wörtl. wir-ergativ verkaufen wir-Dat.-haben-pl.dir.Obj.-ihr-fem.sg.-kann-Prät. = wir hätten sie (pl.) an sie (f.) verkaufen können). Besonders schwierig sind auch die Verben grammatisch-semantisch zu analysieren, da sie oft sehr polysem sind. So drückt das deutsche Präsens die Gegenwart, die Außerzeitlichkeit, die Gewohnheit, die Gesetzlichkeit, das reale Futur, historisches Erlebnis, Aufforderung, reale Bedingung u.a. aus. Lexische und grammatische Bedeutung gehen oft Hand in Hand, z.B. bei transitiven Verben (Valenz), bei Aspektsystemen oder Aktionsarten, beim Ausdruck der verschiedenen Modalitäten. Lexeme können zugleich auch als Grammeme fungieren, z.B. das französische Immediat-Präteritum *je viens de manger* (ich habe soeben gegessen), wo *venir* nicht mehr die Bedeutung 'kommen' hat, die es sonst ausdrückt. Bei Verben der Emotion steht in den kaukasischen Sprachen das Subjekt im Objektskasus, im Finnischen das Objekt im sogenannten Partitiv: *rakastan Teitä* (ich liebe Sie, eigentlich ich liebe einen Teil von Ihnen).

Zur Inhaltserkennung gehört aber vor allem die richtige semantische Analyse aller Lexeme. Die Definition der Sememe kann nicht verbal eingegeben werden, sondern muß digitalisiert werden. Die Digitalisierung muß zugleich systematisch sein, d.h. daß die einzelnen Ziffern Information für Entscheidungsoperationen liefern. Dadurch entsteht die Notwendigkeit einen sprachbezogenen Realienkatalog auszuarbeiten. Sprachbezogen heißt hier, daß nur die in allen der zu verarbeitenden Sprachen befindlichen Sememe katalogisiert werden müssen. Die wichtigsten Noeme (intersprachliche Bedeutungsmerkmale, Begriffselemente) müssen an erster Stelle der Zahl stehen, d.h. daß sie in dem binär-hierarchisch aufgebauten System (Noematikon) die obersten Knoten besetzen. Solche Hauptkategorien sind: Ding vs. Erscheinung, nur-Lebewesen/ nur Nichtlebendes/ Vereinigung von Belebtem und Unbelebtem (Organisation, Institution); die Erscheinungen müssen wieder nach einstelligen und mehrstelligen Prädikaten und diese wieder nach Statik vs Dynamik unterteilt werden. Jedes Semem hat somit eine eigene Zahl, die es im System festlegt. So beginnen alle Lebewesen mit 1, alle Personen mit 11, alle ethnischen Einheiten 111, alle Berufe mit 112, alle Nichtlebewesen (Dinge) mit 2, Kosmika mit 21, alle Institutionen mit 31, alle mehrstelligen Prädikate (Erscheinungen) mit 4, wenn sie statisch sind, mit 5 wenn sie dynamisch sind, Energieübertra-

gung mit 51, Informationsübertragung mit 511, Informationsübertragung mit Effekt mit 5111 usw. Über die Anwendung und den Aufbau des Noematikons werden wir in einem späteren Aufsatz berichten.

Die "Definition" der Sememe muß vorher so vollständig als möglich erfolgen, d.h. es dürfen keine wesentlichen Elemente fehlen, da sie alle zur Monosemierung und zur Beantwortung im System des Mensch-Maschine-Dialogs erforderlich sind. Wir wollen das am deutschen Verbum *widerlegen* darstellen (s. Bild 3). Wir versuchten mit relativ wenigen Leerstellenkriterien auszukommen. Die hier verwendeten sind: S = Energiesender, Y = Energieempfänger, R_i = Relation mit Definition im Index, C = Kommunikationsinhalt, E = Effekt, Π = einstelliges Prädikat; andere Kriterien treffen hier nicht zu.

Widerlegen$_1$

Nicht-Körper ⊃ Energieübertragung ⊃ Informationsübertragung ⊃ Effektive Informationsübertragung ⊃ def. S {homo v collectiv v Produkt ⊃ graphisch} +
+ Y {(homo v collectiv} ⊃ $R_{sapions}$ (Y, Sachverhalt v Hypoth.)}
+ C {Πfalsch (Sachverhalt v Hypothese) + Πargum. [R_{contra} (Beweis, Sachverhalt v Hypothese)]}
+ E {R_{fid} (Y, S)} .

zu lesen: Info-Überträger S (Mensch v Gruppe v Buch) gibt an Info-Empfänger Y (Mensch v Gruppe), die einen Sachverhalt oder Hypothese kennen, eine Information C, beinhaltend falsch-Erklärung des Sachverh. v Hypoth. und Gegenargument mit Beweis gegen Hypothese; mit dem Effekt, daß Y dem S die Beweisführung glaubt.

Bild 3

Der Prozeßablauf der semantischen Analyse erfolgt nun wie folgt: Zunächst wird der Text eingelesen, die Textwörter abgezählt (im allgemeinen die Lexeme bis zum Punkt), Zahlen und Symbole, die im Text verwendet werden, müssen vorläufig abgespeichert werden. Die Lexeme werden sofort dem permanenten Lexikon zugeordnet (Aufsucheprozeß), nicht auffindbare Graphemketten (z.B. Eigennamen) werden zwischengespeichert, aus der Interpunktion wird die Satzart ermittelt und Gliedsätze getrennt (mit Hilfe weiterer Regeln, wie Anzahl der Verben u.a.). Im Lexikon werden die potentiell festen Syntagmen ermittelt (Streckformen, Wortgruppen Spaltverben, Phraseologismen usw.). Aufgrund des grammatischen Lexikoneintrags werden zunächst Wortklasse (Wortart) und Satzgliedgruppen erfaßt und das Prädikativ (Prädikatsverb, Prädikatsqualitativ) festgestellt. Aus dem Lexikon oder speziellen Valenzwörterbuch wird die Valenz des Prädikativs entnommen. Da noch keine Monosemierung stattfand, müssen zunächst alle Valenzen aller Sememe bereitgestellt werden. Aus dem Lexikon werden nun noch alle Grammeme und asemantischen grammatischen Funktionen, wie Kongruenzregeln, Genus, Klassen u.ä. entnommen, wobei die grammatische Kategorie zunächst mit allen Grammemen (d.h. allen Bedeutungen einer grammatischen

Kategorie) bereitgestellt werden muß. Nach Distributionsregeln werden die Grammeme monosemiert und desgleichen werden nach Regeln der Monosemierung, die bei jedem Lexem (Lexemstamm) im Lexikon eingetragen sind, das aktuelle Semem ermittelt, soweit dies innerhalb eines Satzes möglich ist. Das aktuell zutreffende Grammembündel wird zwischengespeichert. Hierauf erfolgt die Sememermittlung bei allen übrigen Lexemen des Satzes ebenfalls nach den Monosemierungsalgorithmen dieser Einheiten. Aufgrund des monosemierten Lexembestands wird die aktuelle Valenz festgestellt, insbesondere die des Prädikativs und als Leerstellenproposition dargestellt. Die Leerstellen werden durch die monosemierten Aktanten als Argumente eingesetzt, so daß die Kernproposition gebildet ist, zu der Satzerweiterungsgruppen, Modalpartikeln, Satzeinleitungswörter und die Gliedsatzkomplexe – hinzugefügt werden. Jetzt können die zwischengespeicherten Teile und Grammeme abgerufen werden und die Proposition kann gebildet werden. Alle Sememe werden in die Notation des Noematikons nach erfolgter Monosemierung transformiert. Die Operationsbefehle – mit 0 beginnende Zahlen – ordnen die Propositionen nach festen Ordnungsregeln so an, daß sie auf die entsprechenden permanent gespeicherten Relationen abgebildet werden können, so daß nur die Argumente die Neuinformation darstellen. Die Recherche-Frage muß so aufgebaut sein, daß sie dieselben Propositionen benutzt und das gefragte Argumentenbündel einer bestimmten Leerstelle durch das entsprechende Fragewort eruiert wird. Wir werden in einem folgenden Aufsatz diesen letztgenannten Mechanismus ausführlicher darstellen, soweit das erarbeitete Modell dies schon zuläßt. Es haben sich dabei eine Menge von neuen Problemen ergeben, die auf bestimmte Mängel verwiesen haben, so z.B. die Auffindbarkeit von Nebeneigenschaften, die nicht für die Sememdefinition wichtig waren, aber unter bestimmten praktischen Gesichtspunkten vordergründig werden können, so z.B. Wertangaben bei Produkten, zeitweilige Eigenschaft, wie Qualitätsminderungen, Absatzkriterien, Leitungskriterien, Wachstumskriterien u.ä. Die semantische Analyse ist zwar das A und O jeder Faktenrecherche, aber sie ist auch der schwierigste Forschungsgegenstand, ganz gleich ob man intralingual oder interlingual herangeht. Da Bedeutungen nicht streng abgezirkelte, logisch streng beschreibbare Größen sind, sondern vielmehr immer offene Radikale nach Kontextelementen zu sein scheinen, ja selbst in der Technik, so besonders in der Rechentechnik, die Verwendung von Metaphern Überhand nimmt (die Maschinen erkennen, suchen, wissen, provozieren, verbergen, verstehen, korrigieren usw.), ist für jedes Neosemem auch ein entsprechender Kontext einzuspeichern.

Schrifftum

Martynov, V.: Univerzal'nyi semantičeskij kod, Minsk, 1977.
Meier, G.F.: Zeichen und Systeme der Sprache, Bd. III, Akademieverlag Berlin, 1966.
Meier, G.F.: Innersprachliche und konfrontative Polysemie – an deutschen und italienischen Beispielen, Zeitschrift für Phonetik XXXI, 1978, pp.143-150.

Eingegangen am 8. November 1983

Anschrift des Verfassers: Prof. Dr. Georg F. Meier, Kavalierstr. 18, DDR-110 Berlin

Problemoj de la semantika analizo por la aŭtomata faktoj-reserĉado
(Resumo)

Komence ni diskutas kelkajn problemojn de lingvistika komputa datumprilaborilo, speciale por la fakt-reserĉo, kiuj baziĝas antaŭ ĉio sur la malperfekta prilaborilo de la semantaj faktoj. Pro tio ke la lingvo ne estas de logika, sed de historia naturo, nur la lingvistika analizo estas eble. Unu de la plej grandaj malfacilaĵoj estas la polisemio de gramatiko kaj de leksiko, ĉar ili postulas multajn komplikitajn decidoperaciojn. Krome ekzistas granda diferencio inter la leksiko-sistemoj de diversaj lingvoj. Ni demonstras ĉi-tion per kelkaj ekzemploj, tiel la esprimoj de la parenceco kaj de la koloroj. Speciale polisemiaj estas la oftaj verboj, ekzemple oni devas traduki la germanan verbon „abnehmen" en la Italan per ĉirkaŭ 40 diversaj verboj; analoge en aliajn lingvojn. Alian malfacilecon ni trovas antaŭ ĉio en la teknika fakvorttresoro; ĉar multaj vortoj estas kombinitaj, sed ili ne estas skribataj entute. Tio-ĉi malfaciligas la identigadon de vortoj. Ankaŭ en la norma lingvo ekzistas multnombraj nepartigeblaj sintagmoj el verbo kaj substantivo. Tiam ni demonstras la monosemigon de polisemiaj gramatikaj kaj leksikaj unuoj, tie-ĉi speciale la monosemigon de pluralaj finaĵoj en la Germana. La kategorioj de subjektivo kaj de objektivo ofte malfacile indentiĝas, tial ke ili estas enkompleksigataj en la formo de la verbo; antaŭ ĉio en la kazo de la ergativo-konstruado. Por stori la signifojn oni bezonas digitalajn unuojn, kiuj devas havi sisteman aranĝon. Tia sistemo estas la „noematikon-o" de nia esplorado kies la aranĝo estas montrata. Per la ekzemplo de germana „widerlegen" (refuti) ni demonstras la metodon de difino de signifo. Fine ni klarigas la procezo -iradon de la enhavo-analizo.

grkg/Humankybernetik
Band 24 · Heft 4 (1983)
Gunter Narr Verlag Tübingen

Mesure de la Durée du Présent et du Moment Psychique Individuel en Termes de Vitesse d'Information

De Juan Carlos CARENA Susana LESPINARD (traduction française) M. del R. SOLHAUNE, J.L. FERRETTI, A. PARDAL, J. PLIEGO, R. ZETA, S. FERNANDEZ, L. TAMAGNO et M. RODRIGUEZ, Rosario (RA)

de l'équipe de Pédagogie Cybernétique, Université Nationale de Rosario, Argentine (Coordinateur: Prof. dr. Juan Carlos CARENA)

1. Introduction

Dans un dessin d'Instruction fondé sur la Pédagogie Cybernétique, on considère six «dimensions de l'espace éducationnel», ou «variables pédagogiques»: l'objectif didactique, la sociostructure, la psychostructure, le milieu informationnel, la matière à apprendre et la systématisation méthodologique.

D'après H. Frank (1976), toute valeur que peut prendre chacune des variables pédagogiques est susceptible d'être représentée par un point dans un système de coordonnées dans un espace de six dimensions. Ainsi donc, tout enseignement possible est déterminé par six valeurs de coordonnées.

Toute tâche qui tend à analyser l'un des composants, aura une correspondance avec une valeur plus adéquate de la variable dans le système, et en conséquence avec une plus grande validité dans le sens d'un enseignement possible.

Ce travail prétend faire des apports à l'analyse de la Psychostructure, grâce à l'étude de certaines caractéristiques du sujet. C'est pour cela que nous avons fait une modification dans la première partie du test K.A.I. (Lehrl, 1980): Lecture de Lettres (BuL). Cette variante naît du besoin d'adapter cette partie-là du test aux hispanophones.

Dès le début, les applications faites en Argentine (zone Rosario), ont démontré que le temps de prononciation des lettres qui exigent 1 émission de plusieurs sons (plus d'un) en langue espagnole (ex: «eme»–«jota»–«ese»–), avait une influence sur le diagnostic de la capacité de l'afflux informatif à la conscience. Par conséquent, on a élaboré selon les critères que l'auteur du test original propose, quatre nouvelles séries de lettres. Elles ont été confectionnées en se servant seulement des quatorze lettres qui, en espagnol, sont prononcées d'une seule émission de voix (Voir tableau 1).

On a testé les sujets simultanément avec les séries des deux structures (de base et expérimentale), avec l'intention de comparer les temps de lecture.

On a démontré, d'après les statistiques, que notres séries, adaptées à l'espagnol, permettent de mesurer l'afflux informatif à la conscience d'une façon plus exacte que les séries élaborées à l'origine pour la langue allemande. Cependant, on démontre aussi que cette structure entretient une étroite corrélation avec l'originale, ce qui permet de

la considérer comme une partie valable d'un nouveau K.A.I. adapté à l'Hispanophone, sans perdre pour cela les objectifs de sa mesure.

A. Structure de Base pour la lecture de lettres, du test original.	B. Structure Expérimentale, avec des sons d'une seule émission de voix, élaborée par nous.
1. unrztrfepkbvdsnildmr	1. ivgakepdocgiqeptbauo
2. IPLZMBEOAEHIOAZTLEAV	2. GBOUKDOPICEKTVEQATID
3. mjztfrdsihdoltkdgwri	3. pavtotqbpivbkdguedcg
4. ECXSBTLKEOGFDEAVIMHP	4. GEOTBTDVCIQVPABUCPOK

Tableau 1

2. *Instruments*

Test Rapide d'Intelligence Générale (K.A.I.). Dans la première partie (Lecture de Lettres: BuL), on a ajouté une structure expérimentale de quatre nouvelles séries (Voir Tableau 1). La deuxième partie (Rétention de digits et de lettres) conserve sa structure originale.

3. *Sujets*

D'un échantillon majeur, on a extrait au hasard 100 cas d'étudiants universitaries des deux sexes des Facultés d'Architecture et de Sciences Basiques de l'Université Nationale de Rosario. Leurs âges oscillent entre 18 et 28 ans et leur participation dans les épreuves a été volontaire.

4. *Résultats*

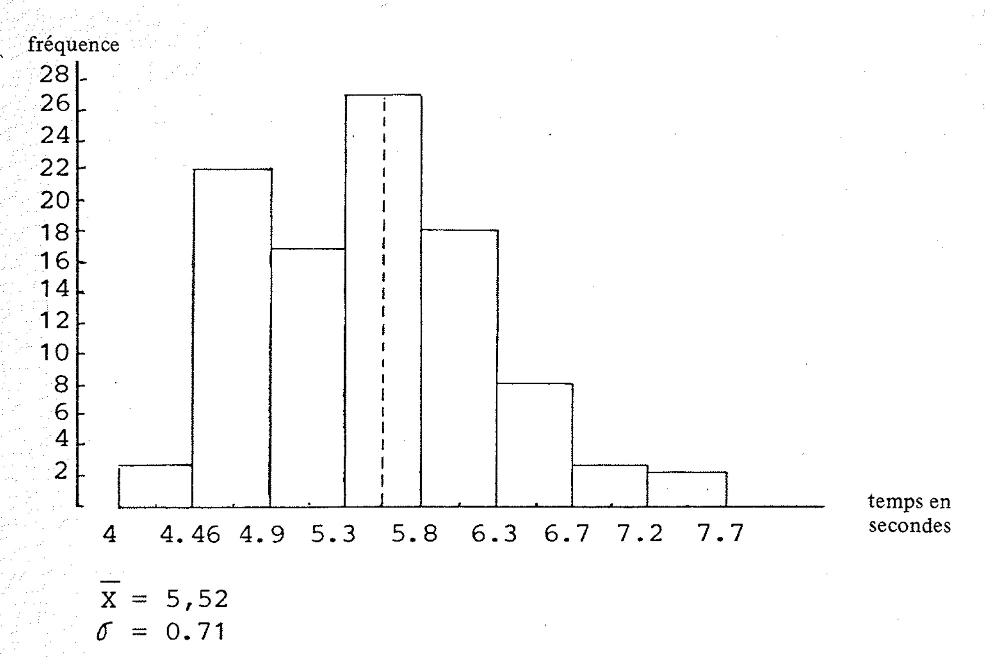
Puisque les résultats des mesures du BuL, dans les structures de Base (A) et Expérimentale (B) s'approchent de la forme d'une courbe normale (Voir graphiques 1 et 2), on présente seulement la table des valeurs du BuL pour les deux structures.

Table No. 1

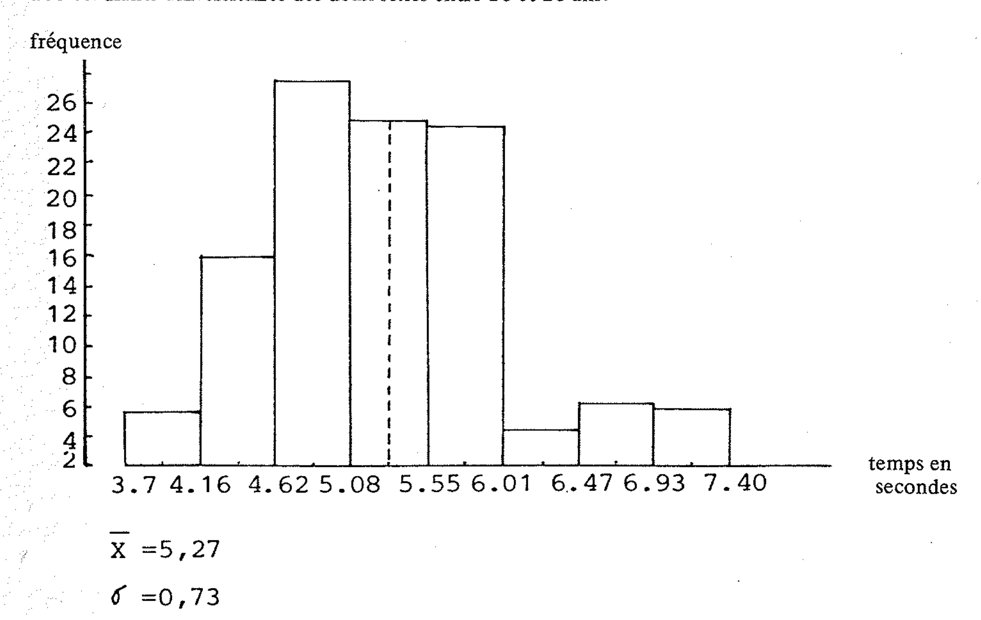
Résultats sur 100 sujets du temps moyen pour la lecture des lettres dans les deux structures.

Structures	Sujets	$\overline{X}$ BuL	σ BuL
De Base (A)	100	5,52	0,71
Expérimentale (B)	100	5,27	0,73

$r = 0{,}81$
$d = 0{,}202$
$Z_d = 4{,}5$ $P < .01$ (Refus de H_0)



Graphique No. 1: Distribution normale des résultats du BuL forme A (de base) obtenus avec 100 étudiants universitaires des deux sexes entre 18 et 28 ans.



Graphique No. 2: Distribution normale des résultats du BuL forme B (expérimentale) obtenus avec 100 étudiants universitaires des deux sexes entre 18 et 28 ans.

Le coefficient de corrélation entre les échantillons: 0,81, entre les deux structures A et B, permet d'assurer la validité des séries proposées pour mesurer le temps de l'afflux informatif à la conscience.

Dans l'application du test paramétrique – pour établir les différences entre les deux variables – on observe une valeur de $Z_d = 4{,}5$, c'est-à-dire que l'on refuse la H_0, et on considère hautement significative, en indiquant que la structure expérimentale demande moins de temps de lecture.

Table No. 2

Lecture des lettres (temps de chaque série)

Sujets		Structure de Base (A)					Structure Expérimentale (B)				
		Série 1	Série 2	Série 3	Série 4	BuL	1	2	3	4	BuL
100	$\bar{x}$	6.41	5.76	6.21	5.95	6.08	6.16	5.61	6.01	5:71	5.87
	S	1.179	0.956	0.918	0.840	1.012	1.068	0.940	0.999	0.909	1.005

F: 1 vs 3: 1.599 $p>0{,}05$
2 vs 4: 1.360 $p>0{,}05$

(1+3) vs (2+4): 16.490 $p<0.001$
F Total: 6.478 $p<0.001$ $F = 5.05$ $p<0.01$

Les résultats de chaque série, tant utilisant la structure de base que l'expérimentale, ont montré une répétition relativement élevée et homogène (coefficient de variation fluctuante entre 14,1% et 18,3%). Cela a motivé des différences entre les temps moyens de réponse pour chaque série: ces différences, (quoiqu'elles soient petites), ont été très significatives dans chaque structure; mais l'hétérogénéité a été plus élevée quand on a utilisé la structure de base (F=6,478; $p<0{,}001$) que quand on a employé la structure expérimentale (F=5,050; $p<0{,}01$). En discriminant entre les séries qui correspondent à la structure de base, selon le nombre de lettres qui demandaient plus d'une seule émission de voix, on a pu déterminer que les séries 2 et 4 (6 lettres de cette classe dans chacune) ont denandé respectivement 5,76 ± 0,96 et 5,95 ± 0,84 secondes (F=1,360; $p>0{,}05$), tant que les séries 1 et 3 (10 et 11 lettres avec ces caractéristiques) ont demandé 6,41 ± 1,18 et 6,21 ± 0,92 secondes (F=1,599; $p>0{,}05$). La différence du temps pour les séries 1 et 3 en moyenne par rapport aux séries No. 2 et 4, a étét très significative (F=16,490; $p<0{,}001$), ce qui indique qu'une source importante de l'hétérogénéité de résultats obtenue avec la structure de base s'est due au nombre de lettres qui ont exigé plus d'une seule émission de voix.

5. Conclusions

- Il est valable de substituer, dans le K.A.I., les séries originales pour la lecture de lettres, par les séries expérimentales.

- La série expérimentale permet de mesurer plus sensiblement la variable temps de lecture, avec une précision majeure dans le diagnostic du BuL.
- Nous offrons la possibilité d'adopter la structure B o expérimentale pour toute application à des hispanophones.

Literature

H. Frank/B.S. Meder: Introducción a la pedagogia cibernética. Editorial Troquel, Buenos Aires, 1976. (Traduit de l'allemand)

S. Lehrl: K.A.I. – Kurztest für allgemeine Intelligenz. Vless, Vaterstetten – München, 1980 (Traduction espagnole du Département des Langues Modernes, Universidad Nacionale de Rosario, 1982).

Reçu 1983-09-03

Adresse des auteurs:
Prof. Dr. Juan Carlos Carena, Grupo Pedagogia cibernetica, Universidad Nacional de Rosario, Riobamba 250 bis, RA-2000 Rosario

Mezuro de la nundaŭro kaj de la subjektiva tempokvanto surbaze de informfluo (Resumo)

Tiu ĉi laboro konsistas el la esplorado de la inteligenteco pere de la aplikado de la testo KAI.
Oni elektis aron da universitataj gestudentoj, kies aĝoj variis inter 18 kaj 28 jaroj.
Oni intencis elpruvi ĉu la literserio komponanta la BUL-n de la originala verko, povis esti anstataŭita de alia literserio adap tita al hispanparolantoj. Hispanlingve, la prononcado de literoj per nura voĉdissendo permesas atingi rezultojn laŭ tiuj starigitaj ĉe la aritmoj de la aŭtoro ellaboritaj.
Oni daŭrigas esploranta la interseriajn variancojn kaj la eblecon eltrovi la incidon de aliaj faktoroj en tiaj rezultoj.

Raporto de la Iniciatgrupo AIS de Eŭropa Klubo pri la plenumita preparlaboro cele la starigon de la Akademio Internacia de la Sciencoj (AIS) Sanmarino

1. Komenco kaj evoluo

La projekto AIS evoluis surbaze de la letero de prof.d-ro Frank de 1981-08-18 al profesorino Marina Michelotti (sekve al pli fruaj proponoj liaj skribitaj al San Marino jam du jarojn antaŭe). Tiun ĉi leteron profesorino Michelotti submetis en itala traduko al la ministrino por klerigado kaj kulturo, d-rino Fausta Morganti. Ekestis pri la propono unua interkonsiliĝo en San Marino la merkredon, 1981-11-25 inter d-rino Morganti, profesorino Michelotti, prof. d-ro Pennacchietti kaj prof.d-ro Frank.

Dua renkontiĝo en RSM kun d-rino Morganti okazis 1982-04-14. Ĝin partoprenis flanke de Eŭropa Klubo la gesinjoroj Argentino, Bozzini, Fantini, Formizzi, H.Frank, B.Frank-Böhringer, I.Frank, Marina Michelotti, Myriam Michelotti, Pennacchietti kaj Raffo. Ili verkis surloke detalan projektproponon kaj transdonis ĝin en ILo kaj la Itala al d-rino Morganti.

Post la akcepto de la propono far la ŝtata kongreso de RSM 1983-05-19 denove okazis interkonsiliĝo kun d-rino Fausta Morganti (1983-07-11), pri kies rezulto ekzistas la kunmetita* protokolo parte diskonigita itallingve per la n-ro 44 de NOTIZZIE STAMPA. Flanke de Eŭropa Klubo partoprenis ĉi tiun kvazaŭ „fondokunsidon“ de la AIS la gesinjoroj Ciccanti, Formizzi, Frank, Marina Michelotti, Pennacchietti kaj Terruzzi. Ne okazis la alvoko de la membroj de fondokomitato per leteroj, kiel antaŭvidite kaj menciite en la protokolo, sed prof.d-ro Frank transprenis la buŝe al li konfiditan taskon, kunlabore kun eksterlandaj profesoroj verki proponojn por statuto kaj ekzamenregularoj, kaj serĉi alvokindajn kaj kunlaborpretajn, elstarajn sciencistojn. La taskoj estas en kunlaboro kun la Iniciatgrupo AIS de Eŭropa Klubo plenumita je la interkonsentita limdato 1983-10-31.

Okazis en San Marino de la 29a ĝis la 31a de oktobro 1983 denova renkontiĝo de la Iniciatgrupo de Eŭropa Klubo por oficiale fini la akceptitan taskon kaj por detale prepari la 1-an Sanmarinan Universitatan Seancon. Tiun kunsidon partoprenis la gesinjoroj Alberto kaj Marinella Balsimelli, Guido, Marina kaj Mŷriam Michelotti, Piera Federici Raffo, Serenella Terruzzi, Argentino, Ciccanti, Fantini, Formizzi, Frank, Pagliarani, Pennacchietti kaj Sammarini.

La laborlingvoj de ĉiuj kunsidoj kaj de la koncerna korespondado estis la Itala kaj la Internacia kun reciproka tradukado.

2. Membroj

Estas rigardendaj kiel membroj de la Iniciatgrupo AIS de Eŭropa Klubo ĉiuj EKanoj, kiuj partoprenis aktive almenaŭ unu de la menciitaj kvar laborkunsidoj en RSM cele la starigon de AIS, t.e. Salvatore Argentino, Verona (I), Alberta Gherardi Balsimelli (RSM), Marinella Balsimelli (RSM), Fiorenza Bozzini, Verona (I), Albino Ciccanti, Rimini (I), Rino Fantini, Cesena (I), Piera Federici Raffo, Chiavari (I), Giordano Formizzi, Verona (I), Helmar Frank, Paderborn (D), Ines Frank, Paderborn (D), Brigitte Frank-Böhringer, Paderborn (D), Duilio Magnani, Rimini (I), Guido Michelotti (RSM), Maria Luisa Michelotti (RSM), Marina Michelotti (RSM), Myriam Michelotti (RSM), Romeo Pagliarani, Cesena (I), Fabrizio Pennacchietti, Torino (I) kaj Serenella Giacchino Terruzi, Milano (I). - Krome prof.d-ro Frank havigis al si - kiel interkonsentite kun d-rino Fausta Morganti - la apogon de diversaj eksterlandaj sciencistoj (preskaŭ senescepte EKanoj) por starigi statut- kaj regularproponojn.

3. Rezultoj

Al la ministrino por klerigado kaj kulturo la Iniciatgrupo povas ek de hodiaŭ transdoni
(a) statutproponon por AIS
(b) proponojn por la ekzamenregularoj
(c) interkonsentitan liston de alvokindaj kaj kunlaborpretaj sciencistoj
(ĉ) rezolucion kun rekomendoj pri la estonta procedo**
(d) ĉi raporton.

La Iniciatgrupo, plenuminte ĉi taskaron, plulaboros
- preparante la 1-an Sanmarinan Universitatan Seancon anoncitan per la n-ro 44 de Notizie Stampa por la tempo post kristnasko 1983
- restante je dispono de la ministrino por klerigado kaj kulturo por eventualaj pluaj interkonsiliĝoj aŭ taskoj.

San Marino, 1983/1683 pfR - 10 - 31.
(Subskriboj de ĉiuj ĉeestaj Iniciatgrupanoj)

* vd. grkg/Humankybernetik 2/83, p.93-94
** ĉi sekvas

Rezolucio kaj rekomendoj de la Iniciatgrupo de Eŭropa Klubo por la starigo de AIS
San Marino, 1983/1683pfR - 10 - 29

La iniciatgrupo AIS de Eŭropa Klubo kunveninta en San Marino de la 29a ĝis la 31a de oktobro 1983/1683pfR decidis la jenan rezolucion kaj rekomendojn al la Ministrino pri Klerigado kaj Kulturo, d-rino Fausta MORGANTI:

1) Ni konstatas, ke nek venis la iniciato por la fondo de la Akademio Internacia de la Sciencoj el la Esperanto-Movado, nek kontribuis ĝis nun iu Esperantista organizaĵo al la starigo de AiS, nek estu celo de AIS konkurenci kun la Akademio de Esperanto aŭ varbi por la Internacia Lingvo. Ĉi tiu lingvo (krome akceptata, ĉar efika por modernaj informiloj, far kibernetikistoj) nur servas kiel unu el la oficialaj kaj laboraj lin-

(daŭrigo: paĝo 192)

grkg/Humankybernetik
Band 24 · Heft 4 (1983)
Gunter Narr Verlag Tübingen

Ein Algorithmus zur Ähnlichkeitsuntersuchung deutscher Vornamen

Algoritmo por decido pri simileco de germanaj antaŭnomoj

von / de

Rudolf-Josef FISCHER, Münster (D)

aus dem / el la
Institut für Medizinische Informatik und Biomathematik der Westfälischen Wilhelms-Universität Münster
(Direktor: Prof. Dr. F. Wingert)

1. *Problemstellung*

Täglich werden Hunderte von Patienten in den Universitätskliniken Münster aufgenommen. Bei jeder Aufnahme muß kontrolliert werden, ob zu dem Patienten schon Daten früherer Aufnahmen vorliegen. Diese Überprüfung ist Aufgabe der „automatischen Identifizierung“, die sich auch in vielen anderen Bereichen (Einwohnermeldeamt, Kundenkartei, usw.) ergibt. Für die Identifizierung werden Angaben wie Geburtsdatum, Name, Vorname, Geschlecht und Geburtsort verwendet.

Dabei muß berücksichtigt werden, daß diese Angaben fehlerhaft oder unvollständig sein können. Oft gibt gerade ein Vergleich der Vornamen den Ausschlag, ob das Heranziehen weiterer Daten noch Erfolg verspricht. Dafür muß die Ähnlichkeit zweier Vornamen automatisch bestimmt werden können.

1. *Enkonduko*

Ĉiutage centoj da pacientoj estas akceptataj en la universitata klinikaro de Münster. Dum ĉiu akcepta proceduro oni devas kontroli, ĉu por tiu paciento jam haveblas datumoj de antaŭaj restadoj. Tiun kontrolon faras la „aŭtomata identigo“, kiun oni ankaŭ aplikas sur multaj aliaj kampoj (enloĝantara administrejo, klienta adresaro, ktp.) Por la identigo oni uzas informojn kiel naskiĝodaton, nomon, antaŭnomon, sekson kaj naskiĝlokon.

Ĉe tio oni devas konsideri, ke tiuj informoj povas esti erarenhavaj aŭ malkompletaj. Ofte ĝuste la komparo de la antaŭnomoj decidas, ĉu la uzo de pliaj datumoj ankoraŭ estas sukcespromesa. Pro tio la simileco de du antaŭnomoj devas esti aŭtomate kalkulebla.

2. *Der Begriff „Ähnlichkeit“ bei Vornamen*

Im Gegensatz zum Familiennamen kommt es bei Vornamen häufig vor, daß

2. *La nocio „simileco“ de antaŭnomoj*

Kontraŭe al familia nomo homoj ofte uzas ne la oficiale registritan formon de sia antaŭnomo, sed varianton. Oni devus

Anmerkung: Ausnahmsweise wird hier ein Beitrag zweisprachig abgedruckt. Er diene als Muster für wissenschaftliche Texte (Magister-, Dissertations- oder Habilitationsarbeiten) an der Internationalen Akademie der Wissenschaften San Marino.

nicht die amtlich registrierte Form, sondern eine Variante geführt wird. Es müßten also zu einem Vornamen alle seine Varianten als ähnlich erkannt werden können. Diese Varianten hängen stark von der Landessprache ab, zum Teil sogar mit regionalen Unterschieden. Deshalb ist der hier vorgestellte Algorithmus im wesentlichen auf deutsche Vornamen zugeschnitten, obwohl er auch bei ausländischen Vornamen oft ein richtiges Ergebnis liefert.

do aŭtomate rekoni ĉiujn variantojn de iu antaŭnomo kiel similaj al ĝi. Tiuj variantoj tre dependas de la landa lingvo, parte eĉ de regiona. Tial la suba algoritmo taŭgas ĉefe nur por germanaj antaŭnomoj, kvankam ĝi liveras ankaŭ pri alilandaj antaŭnomoj ofte la ĝustan rezulton.

Zwei vorgegebene Vornamen können einen Ähnlichkeitsgrad von 0 bis 3 haben. 0 bedeutet „keine Ähnlichkeit“, 1 bis 3 geben zunehmende Ähnlichkeit an. Die Definition des Begriffes „Ähnlichkeit“ bei Vornamen ergibt sich aus der Beschreibung des Ähnlichkeitsgrades.

Du donitaj antaŭnomoj povas havi similecogradon de 0 ĝis 3. 0 signifas malsimilecon, 1 ĝis 3 kreskantan similecon. La difino de „simileco“ de antaŭnomoj rezultas el la algoritmo, kiu kalkulas la similecogradon.

Allgemein gilt, daß zwei identische Vornamen zwar den höchsten Ähnlichkeitsgrad 3 erhalten, daß aber die Berechnung nicht besonders schnell ist, da man davon ausgehen sollte, daß nur bei zwei nicht identischen Vornamen eine Ähnlichkeit berechnet wird.

Ĝenerale oni premisu, ke oni uzas la algoritmon nur por malidentaj antaŭnomoj. Identaj kompreneble rezultigas similecogradon 3, sed la kalkulo ne estas specife rapida.

Ferner gilt eine durch Schreibfehler entstandene Variante nicht als ähnlich. Eine durch Schreibfehler definierte Ähnlichkeit läßt sich mit dem in (1) beschriebenen Algorithmus berechnen.

Plie, varianto, kiu ekiĝis per skrib-eraro, ĉi tie ne rigardatas kiel simila. Skrib-erare difinitan similecon oni povas kalkuli per algoritmo el (1).

3. *Kriterien für den Ähnlichkeitsgrad 3*

3. *Kriterioj por la similecogrado 3*

Im folgenden werden zunächst Transformationen beschrieben, mit denen Vornamen zur Überprüfung von Ähnlichkeitsgrad 3 verändert werden. Zu jeder Transformation wird ein ähnliches Vornamenspaar als Begründung angegeben.

Unue mi sube priskribos kelkajn transformojn, kiuj modifas antaŭnomojn por decidi pri similecogrado 3. Por ĉiu transformo mi prezentas motivantan paron de similaj antaŭnomoj.

a) Streichen aller Leerzeichen und Bindestriche
a) Forigo de ĉiuj interspacoj kaj streketoj

Beispiel: / Ekzemplo:

KARL HEINZ, KARLHEINZ

und / kaj

KARL-HEINZ

sind ähnlich. / estas similaj.

b) TH wird durch T ersetzt
b) TH anstataŭatas per T

Beispiel: / Ekzemplo:

GUENTHER GUENTER

c) PH wird durch F ersetzt
c) PH anstataŭatas per F

Beispiel: / Ekezmplo:

STEPHAN STEFAN

d) CH wird durch C ersetzt
d) CH anstataŭatas per C

Beispiel: / Ekzemplo:

ERICH ERIC

e) K wird durch C ersetzt
e) K anstataŭatas per C

Beispiel: / Ekzemplo:

MARK MARC

f) Z wird durch C ersetzt
f) Z anstataŭatas per C

Beispiel: / Ekzemplo:

FRANZISKA FRANCISKA

g) -L, -ER, -EL oder -CHEN am Ende wird gestrichen
g) -L, -ER, -EL aŭ -CHEN finaj malaperas

Beispiele: / Ekzemploj:

HEIN HEINER

HANS, HANSEL

und / kaj

HANSL

FRITZ FRITZCHEN

Ist nach diesen Transformationen einer der Vornamen im anderen als Teilzeichenreihe enthalten, haben die beiden gegebenen Vornamen den Ähnlichkeitsgrad 3.

Se post tiuj transformoj unu el la du antaŭnomoj estas subsignovico de la alia, la donitaj antaŭnomoj havas la similecogradon 3.

Man beachte, daß die obigen Transformationen keineswegs immer wie in den Beispielen zu einer gültigen Variante führen müssen. Entscheidend ist, daß der wichtigste Wortteil übrig bleibt, der dann evtl. als Teilzeichenreihe in dem anderen Vornamen vorkommt.

Oni atentu, ke la supraj transformoj ne ĉiam devas rezultigi validan varianton kiel en la ekzemploj. Ĉefe gravas, ke plej tipa nomparto restas, enhavate eble kiel subsignovico en la alia antaŭnomo.

Beispiel: / Ekzemplo:

Gegeben seien / Donitaj estu

KAETHE KAETCHEN

Die Transformationen führen zu: / La transformoj rezultigas:

KAETE KAET

KAET ist Teilzeichenreihe von KAETE. Also haben KAETHE und KAETCHEN den Ähnlichkeitsgrad 3.

KAET estas subsignovico de KAETE. Tial KAETHE kaj KAETCHEN havas similecogradon 3.

4. *Kriterien für die Ähnlichkeitsgrade 2 und 0*

4. *Kriterioj por la similecogradoj 2 kaj 0*

Falls zwei gegebene Vornamen nicht den Ähnlichkeitsgrad 3 erreichen, werden sie weiter transformiert.

Se du donitaj antaŭnomoj ne havas la similecogradon 3, ili plu transformatas.

a) Ein Vokal oder ein Y am Ende wird gestrichen
b) AE wird durch A ersetzt
c) OE wird durch O ersetzt
d) In Vornamen, die nach diesen Transformationen noch mehr als 3 Buchstaben haben, werden aufeinanderfolgende Zeichen bis auf eines gestrichen.

a) Fina vokalo aŭ Y malaperas
b) AE anstataŭas per A
c) OE anstataŭas per O
d) Se antaŭnomoj post tiuj transformoj ankoraŭ havas pli ol 3 literojn, el ĉiuj apudaj identaj literoj restas nur unu.

Falls die beiden Vornamen nach diesen Schritten weiterhin ungleich sind und einer von ihnen weniger als 3 Buchstaben lang ist, werden die ursprünglichen Vornamen als unähnlich festgestellt (Ähnlichkeitsgrad 0). Wenn sie jedoch beide noch mindestens 3 Zeichen lang sind und eine von ihnen in der anderen als Teilzeichenreihe enthalten ist, haben die ursprünglichen Vornamen den Ähnlichkeitsgrad 2.

Se la du antaŭnomoj post tiuj proceduroj plu estas malidentaj kaj se unu el ili nun havas malpli ol 3 literojn, la originaj antaŭnomoj estas malsimilaj (similecogrado 0). Sed se ambaŭ ankoraŭ havas minimume 3 literojn kaj se unu el ili estas nun subsignovico de la alia, la originaj antaŭnomoj havas la similecogradon 2.

Beispiele: / Ekzemploj:

1) AENNCHEN ANNA

Transformation 3g führt zu: / Transformo 3g rezultigas:

AENN ANNA

Kein Ähnlichkeitsgrad 3! / Ne similecogrado 3!

4a: AENN ANN

4b: ANN ANN

4d unterbleibt, da nur noch 3 Zeichen übrig sind. Weil die transformierten Vornamen nun gleich sind, haben die ursprünglichen den Ähnlichkeitsgrad 2.

4d ne aplikatas, ĉar restis nur 3 literoj. Ĉar la transformitaj antaŭnomoj estas nun identaj, la originaj havas la similecogradon 2.

2) HELMUTH HELLMUT

3b führt zu: / 3b rezultigas:

HELMUT HELLMUT

4d: HELMUT HELMUT

Es ergibt sich also Ähnlichkeitsgrad 2. / Rezultas do similecogrado 2.

3) KAY KAI

4a führt zu: / 4a rezultigas:

KA KA

Die beiden Vornamen haben nun zwar weniger als 3 Zeichen, sind aber identisch, so daß sie den Ähnlichkeitsgrad 2 erhalten.

La du antaŭnomoj jes nun havas malpli ol 3 literojn, sed identas, kaj tial ricevas la similecogradon 2.

4) JOCKEL JOCHEN

3e und 3d führen zu: / 3e kaj 3d rezultigas:

JOCCEL JOCEN

3g führt zu: / 3g rezultigas:

JOCC JOCEN

4d: JOC JOCEN

JOCKEL und JOCHEN erhalten den Ähnlichkeitsgrad 2, da JOC als Teilzeichenreihe in JOCEN enthalten ist.

JOCKEL kaj JOCHEN ricevas la similecogradon 2, ĉar JOC estas subsignovico de JOCEN.

5. *Kriterien for die Ähnlichkeitsgrade 1 und 0*

5. *Kriterioj por la similecogradoj 1 kaj 0*

Falls die transformierten Formen zweier Vornamen auch nicht den Ähnlichkeitsgrad 2 haben, wird geprüft, ob die Zeichen der kürzeren in der längeren in derselben Reihenfolge vorkommen. Falls ja, haben die ursprünglichen Vornamen den Ähnlichkeitsgrad 1; sonst sind sie unähnlich (Ähnlichkeitsgrad 0).

Se la transformitaj formo de du antaŭnomoj ankaŭ ne havas la similecogradon 2, kontrolatas, ĉu la literoj de la pli mallonga estas samvice en la pli longa. Se jes, la originaj antaŭnomoj havas la similecogradon 1; alikaze ili estas malsimilaj (similecogrado 0).

Beispiele: / Ekzemploj:

1) HEINRICH HEINZ

3d und 3f führen zu: / 3d kaj 3f rezultigas:

HEINRICH HEINC

Die Zeichen von HEINC kommen in derselben Reihenfolge in HENRIC vor; also Ähnlichkeitsgrad 1.

La literoj de HEINC aperas samvice en HEINRIC; do similecogrado 1.

2) FRIEDRICH FRITZ

3d und 3f führen zu: / 3d kaj 3f rezultigas:

FRIEDRIC FRITC

Die Zeichen von FRITC kommen nicht alle, nämlich das T nicht, in FRIEDRIC vor; also Unähnlichkeit, obwohl FRITZ eine Variante von FRIEDRICH ist.

El la literoj de FRITC ne ĉiuj, nome ne la T, estas en FRIEDRIC; do malsimileco, kvankam FRITZ estas varianto de FRIEDRICH.

6. *Genauigkeit des Algorithmus*

6. *Precizeco de la algoritmo*

Natürlich kann man nicht erwarten, daß der Algorithmus in allen Fällen genauso

Kompreneble oni ne povas atendi, ke la algoritmo en ĉiuj okazoj same decidas pri

auf Ähnlichkeit entscheidet wie ein Mensch. Selbst für den Menschen gibt es Paare von Vornamen, bei denen die Meinungen auseinandergehen, ob der eine Vorname eine Variante des anderen ist. Der oben beschriebene Algorithmus geht davon aus, daß Varianten durch Erweitern oder Abkürzen einer Grundform entstehen.

Männliche und weibliche Variante derselben Grundform werden als ähnlich angesehen, weil, wie die Praxis zeigt, aus einem fehlerhaft notierten Vornamen oft ein falsches Geschlecht abgeleitet wird. Die zweifache Berücksichtigung einer Abweichung im Geschlecht könnte eine zu geringe Wahrscheinlichkeit auf Identität zweier Personen vortäuschen und so eine Strategie der automatischen Identifizierung zum Scheitern bringen.

Bisher liegt eine nur kleine Stichprobe von Vornamenspaaren vor, die sich aus der Patientenaufnahme mehrerer Universitätskliniken in Münster ergab.

Insgesamt handelt es sich um 119 Fälle (siehe Tabelle 1), in denen zu derselben Person ein abweichendes Vornamenpaar vorkam. Da der Ähnlichkeitsalgorithmus auf Varianten deutscher Vornamen zugeschnitten ist, müssen zunächst 16 Fälle von ausländischen Vornamen und 14 Fälle von Schreibfehlern ausgeklammert werden. Von diesen erkannte der Algorithmus aber immerhin noch jeweils 7 Paare als ähnlich. Ferner kam 1 sonstiger Fehler, nämlich eine unzulässige Abkürzung („B." für „BARBARA") vor. Es verbleiben also 88 Fälle, die einen Hinweis auf die Genauigkeit des Algorithmus erlauben.

Davon wurden 79 als ähnlich erkannt. In 9 Fällen versagte der Algorithmus, darunter allein viermal bei einem Paar DETLEF und DETLEV. Die Erfolgsquote beträgt also etwa 90%.

simileco kiel homo. Sed eĉ por homoj ekzistas antaŭnomaj paroj, pri kiuj oni malsamopinias, ĉu el la du antaŭnomoj unu estas varianto de la alia. La supre prezentita algoritmo hipotezas, ke variantoj ekiĝas kiel plilongaĵo aŭ kiel malplilongaĵo de baza versio.

Porviraj kaj porvirinaj variantoj de la sama baza versio rigardatas kiel similaj, ĉar, kiel spertigas la praktiko, erara antaŭnomo poste povas konduki al miskorekto de la seks-indiko. La duobla konsidero de malidentaj seksoj povus ŝajnigi maltro probablan identecon de du personoj kaj malsukcesigi strategion de aŭtomata identigo.

Ĝis nun haveblas nur malgranda samplo de antaŭnomaj paroj, kiu rezultis el la akcepta proceduro de pacientoj en pluraj universitataj hospitaloj de Münster.

Estas entute 119 paroj (vidu tabelon 1) de identaj personoj kun malidentaj antaŭnomoj. Ĉar la algoritmo ĉi trovu nur variantojn de germanaj antaŭnomoj, oni unue ne konsideru 16 alilandajn parojn kaj 14 parojn kun skrib-eraro. El ili tamen la algoritmo eltrovis ne malpli ol po 7 similajn. Plie okazis 1 alia eraro, nome malpermesita mallongigo („B." anstataŭ „BARBARA"). Restas do 88 paroj, kies rezultoj vere rilatas al la precizeco de la algoritmo.

El tiuj eltrovatis 79 similaj. 9-foje la algoritmo malsukcesis, el tiom jam 4-foje pro iu paro DETLEF kaj DETLEV. La sukceskvanto do estas ĉirkaŭ 90%.

Vornamen *antaŭnomoj*	Anzahlen *nombroj*	Ähnlichkeitsgrad *similecogrado* >0	=0	nach zusätzlichem Einsatz des Algorithmus nach [1] *post kroma apliko de algoritmo laŭ* [1] ähnlich *similaj*	unähnlich *malsimilaj*
deutsche *germanaj*	Varianten *variantoj*	79	9	87	1
	Schreibfehler *skrib-eraroj*	7	7	13	1
	sonstige Fehler *aliaj eraroj*		1		1
ausländische *alilandaj*	(Varianten oder Fehler) *(variantoj aŭ eraroj)*	7	9	14	2
Summe *sumo*		93	26	114	5

Tabelle 1: Ergebnisse der Ähnlichkeitsuntersuchung von 119 Paaren abweichender Vornamen identischer Personen ohne und mit zusätzlichem Einsatz des Algorithmus aus [1]

tabelo 1: rezultoj de decido pri simileco ĉe 119 paroj da malidentaj antaŭnomoj de identaj personoj, sen kaj kun kroma apliko de la algoritmo laŭ [1]

In drei Fällen meldete der Algorithmus Ähnlichkeit bei abweichenden Vornamen nicht identischer Personen, aber nur in einem dieser Fälle (HANS und *H*ERM*A*N*N*-JO*S*EF) lag offensichtlich ein Versagen der Ähnlichkeitskriterien vor.

Um zwei Personen zu identifizieren, kann man für als unähnlich gemeldete Vornamen zusätzlich einen Algorithmus aus FISCHER (1982) einsetzen, der überprüft, ob die Unähnlichkeit auf Schreibfehlern beruht. Dieser Algorithmus erkannte von

3-foje la algoritmo anoncis similecon de malidentaj antaŭnomoj de malidentaj personoj, sed nur 1-foje (HANS kaj *H*ER*M*A*N*N-JOSEF) tio signifis senduban malfunkcion de la similec-kriterioj.

Por identigi du personojn, oni povas por antaŭnomaj paroj, ĝis tiam kiel malsimilaj konstatitaj, apliki krome algoritmon laŭ FISCHER (1982), kiu kontrolas, ĉu la malsimilecon kaŭzis skrib-eraroj. Tiu algoritmo malkovris, ke inter la restintaj 9

den restlichen 9 ausländischen Vornamenspaaren 7 als ähnlich, obwohl ohne Kenntnis der Bildungsgesetze für Varianten bei ausländischen Vornamen nicht entschieden werden konnte, ob es sich wirklich um Schreibfehler und nicht vielmehr um Varianten handelte. Von den verbliebenen 7 Fällen von Schreibfehlern erkannte das Verfahren aus FISCHER (1982) 6 als ähnlich. Bei dem 7. Fall (KAY und KAZ) war der Vorname KAY zu kurz, als daß man per Algorithmus KAZ als Schreibfehler hätte behaupten können.

Bei zusätzlichem bedingtem Einsatz des Algorithmus aus FISCHER (1982) wurden also insgesamt von den 119 Fällen 2 ausländische Vornamenpaare, 1 Schreibfehler, 1 sonstiger Fehler und nur 1 weiterer Fall (MARLIES und MARIA-ELISABETH) nicht als ähnlich erkannt.

Durch Fortschreiben dieser Statistik und damit durch Vergrößern der Stichprobe wird sich überprüfen lassen, ob das obige Ergebnis schon eine zuverlässige Schätzung der Genauigkeit darstellt. Falls weiterhin Paare wie DETLEF und DETLEV vermehrt auftauchen, könnte dieses noch zu einem Abändern der Kriterien für die Ähnlichkeitsgrade führen.

alilandaj antaŭnomaj paroj estis 7 similaj, kvankam oni sen scio pri la ekiĝ-reguloj por variantoj de alilandaj antaŭnomoj ne povas decidi, ĉu temis fakte pri skriberaroj aŭ kontraŭe pri variantoj. El la restintaj 7 paroj kun skrib-eraroj la algoritmo el FISCHER (1982) ekkonis, ke 6 estis similaj. La nomoj de la 7-a paro (KAY kaj KAZ) estis tro mallongaj por algoritme aserti, ke KAZ estas skrib-eraro de KAY.

Do, post kroma laŭbezona apliko de la algoritmo laŭ FISCHER (1982) el la 119 paroj entute restis 2 alilandaj, 1 skrib-erara, 1 ali-erara kaj nur 1 lasta (MARLIES kaj MARIA-ELISABETH) laŭalgoritme malsimilaj.

Post daŭrigo de tiu statistika nombrado kaj do per la tiel kreskanta samplo kontroleblos, ĉu la supra rezulto jam estas fidinda stimaĵo de la precizeco. Se paroj kiel DETLEF kaj DETLEV plue aperas sufiĉe ofte, tio povus konduki al ŝanĝo de la similec-kriterioj.

Schrifttum *Literaturo*

FISCHER, Rudolf-Josef: Trovado de similaj vortoj en ampleksa vortaro. En: H.Frank/Yashovardhan/B.Frank-Böhringer (eld.): Lingvokibernetiko, Tübingen 1982, Gunter Narr

Eingegangen am Ricevita je

1983-07-01

Anschrift des Verfassers: Adreso de la aŭtoro:

Dr. Rudolf-Josef Fischer, Institut für Medizinische Informatik und Biomathematik der Westfälischen Wilhelms-Universität Münster, Hüfferstr. 75, D-4400 Münster

An algorithmus for measuring the similarity of German names (summary)

The admission procedure at the Münster University Hospital stipulates that for each newly admitted patient one must check whether any records of previous admissions exist. This check is carried out by means of an algorithmus for "automatic identification" which compares date of birth, surname, given name, sex and possibly place of birth. One must also take into account the fact that some of these data may be incomplete or erroneous. The decision whether to carry on with the search often rests just on the comparison of given names. For this purpose the degree of similarity between given names must be made automatically measurable.
In place of the officially registered form of their given names people often use variants, so all one needs is to automatically "recognise" all variants of a certain name as being similar to its "official" form. For our purpose we define 4 degrees of similarity ranging from 0 to 3. 0 implies non-identity, 1 to 3 increasing degrees of similarity. The definition of similarity is a direct consequence of the algorithm used to measure it; in fact it decreases according to the number of transformations required to acheive identity. These transformations take into consideration the usual ways of abbreviating German names and of attaching dimunitive suffixes to them. In spite of this one acheives good results even when dealing with non-German names.
A test showed a success-rate of about 90% and the additional application of an algorithm fo locating type-errors increased this to approximately 99%.

Offizielle Bekanntmachung

Raporto de la Iniciatgrupo AIS (daŭrigo de paĝo 182)

gvoj de AIS. La Internacia Lingvo tamen estas nemalhavebla komunikilo de AiS pro la neŭtraleco de San Marino kune kun la fakto, ke junaj sciencistoj de orienteŭropaj landoj multe pli facile ol en la Angla aŭ la Franca esprimiĝas en la Rusa lingvo, kiu inverse malofte estas komprenata far la okcidenteŭropaj sciencistoj. - Por ke ne ekestu ajna miskompreno pri la nur rimeda rolo de la Internacia Lingvo en AIS, ni insiste rekomendas eviti en ĉiuj oficialaj tekstoj kaj gazetarinformoj rilate AIS la kromnomon „Esperanto" de la Internacia Lingvo sed uzi nur ĉi tiun originalan nomon, eventuale mallongigite kiel I.L. aŭ ILo.

2) Ni konstatas, ke laŭ interkonsento de 1983-07-11 inter la iniciatgrupo kaj la ministrino pri klerigado kaj kulturo estis verkita 1983-07-27 unua propono de statuto kiel bazo de koncernaj diskutoj kun eksterlandaj universitataj profesoroj, precipe inter profesoro Frank (D) kaj la profesoroj Bociort (R), Ĉen (TJ), Formizzi (I), Haszpra (H), Lapenna (GB), Marinov (BG), Neergaard (DK), Pennacchietti (I), Sangiorgi (BR) , Sherwood (USA), Szerdahelyi (H) kaj Weltner (D), sed - skribe - ankaŭ kun pluraj aliaj sciencistoj. Surbaze de la diversaj ŝanĝproponoj la profesoroj Frank kaj Pennacchietti verkis en ILo kaj la Itala ĝustatempo la promesitan proponon por la Ministrino pri Klerigado kaj Kulturo kun dato 1983/1683pfR-09-26, kompostis kaj storis sur magnetkartojn la proponon por poste ebligi rapidajn ŝanĝojn, kaj tuj presigis kelkdek ekzemplerojn por la ministrino, la membroj de la iniciatgrupo kaj la kunlaborintaj sciencistoj. Publikigo ankoraŭ ne okazis kaj ne okazu antaŭ la permeso far la ministrino. - La iniciatgrupo rekomendas la akcepton de la statutpropono kun la jenaj esceptoj:

a) en §5.1 estu interŝovita inter „kibernetikajn" kaj „kaj la prikulturajn" la klaŭzo „la prinaturajn (precipe la biologiajn kaj ekologiajn)"

b) en §9.3 la klaŭzoj „kolegianoj, inter kiuj minimume du estas" kaj „laŭ elekto de la apartenantoj" estu forstrekitaj. Estu aldonata: „Dekomence funkciu la jenaj tri sekcioj: Kibernetiko (provizore kun filozofio), kultursciencoj (provizore kun morfosciencoj) kaj natursciencoj (provizore kun struktursciencoj)."

Ni rekomendas, tuj post la akcepto far la Ŝtata Kongreso publikigi la statuton Itale (kaj laŭeble ILe) en Civilta' Cibernetica, kaj ILe (kaj laŭeble angle aŭ france) en grkg/Humankybernetik.

3) Ni konstatas, ke ankaŭ por la tri ekzamenregularoj tuj (jam en la julio) estis verkataj unuaj skizoj kiel diskutbazoj, kaj ke bontempe antaŭ la limdato interkonsentita kun la ministrino pri klerigado kaj kulturo la 11-an de julio la skizoj estis transformataj en proponojn transdonotajn al la ministrino. La tekstoj - ĝis nun same kiel en la kazo de la statuto nur etkvante presitaj por la ministrino, la iniciatgrupo kaj la kunlaborintoj - troviĝas sur magnetkartoj, tiel ke ŝanĝoj kaj presigo rapide eblas post akcepto. - Ni rekomendas, ke almenaŭ la regularo por la docentiĝo estu decidata jam kune kun la statuto, por ke AIS povu eklabori. (...)

4) Transprenante la taskon menciitan en §9.2 de la statutpropono kaj aprobinte la kriteriojn starigitajn far la profesoroj Frank kaj Pennacchietti, ni rekomencas al la ministrino pri klerigado kaj kulturo la alvokon de

a) la universitatnivelaj profesoroj sur la listo 4.1(1.1) kiel honoraj profesoroj kaj dumvivaj sciencaj senatanoj

b) la universitatnivelaj profesoroj sur la listo 4.1(2.1) kiel honoraj profesoroj kaj asociitaj membroj

c) la meritplenaj elstudintaj sciencistoj sur la listo 4.1(2.3) kiel adjunktoj

ĉ) la meritplenaj subtenantoj sur la listo 4.1(3) kiel konsilantoj.

5) Ni decidis transpreni la organizan respondecon por la okazigo de 1-a Sanmarina Universitata Seanco (SUS) de la 27a de decembro ĝis la 7a de januaro 1683pfR, en kies kadro povas okazi la Inaŭguro de AIS, se tion volas la registaro de RSM. Ni petas niajn membrojn en San Marino, Sinjorinoj Marina kaj Myriam Michelotti, surlokeпе preni ĉiujn necesajn iniciatojn kaj precipe sendi la necesajn oficialajn invitleterojn al la partoprenontoj. Laŭ la volo de la ministrino pri klerigado kaj kulturo AIS povas lastminute fariĝi la oficiala okaziganto de la tuta 1-a SUS aŭ de parto de ĝi.

6) por ne seniluziigi la elstarajn sciencistojn el la tuta mondo, kiuj entuziasme aprobas AIS, ni samopinias, ke ia konstituiĝo de AIS devos okazi dum la unua SUS, kiel anoncite por la semajno post kristnasko 1983 per la Notizie Stampa n-ro 44 de la 13a de julio 1983. Tamen estas kvar diversaj ebloj plenumi ĉi tiun deziron:

I) Se la Ŝtata Kongreso laŭplane antaŭ la 19a de novembro decidas statuton de AIS kaj rajtigas la Ministrinon pri Klerigado kaj Kulturo alvoki la unuajn membrojn, AIS povos plenforte ekfunkcii per aŭ dum la unua SUS.

II) Se la Ŝtata Kongreso ankoraŭ ne decidas statuton sed ja konkretigas la fondodecidon de la 19a de majo per la rajtigo al la Ministrino pri Klerigado kaj Kulturo alvoki la unuajn membrojn, tiuj ĉi povos kontitui dum la unua SUS la Akademion kaj transpreni la taskon rediskuti kaj - konforme al direktivoj donotaj far la ministrino - revizii niajn statut- kaj regularproponojn.

III) Se la Ŝtata Kongreso ne decidas statuton kaj prokrastas la inaŭguron de AIS al difinita dato, tiam la alvokindaj sciencistoj menciitaj sur la kunmetitaj listoj povos dum la unua SUS konstituiĝi kiel provizora Kolegio de AIS kaj plenumi necesajn preparlaborojn, pli-malpli apogante sin sur nian statutproponon.

IV) Se la Ŝtata Kongreso nek decidas statuton, nek rajtigas la Ministrinon tuj alvoki la unuajn membrojn, nek prokrastas la planitan inaŭguron al difinita dato en la venonta jaro, tiam la sciencistoj sopirantaj al la ekesto de AIS memstare ĝin povos konstitui dum la unua SUS kaj poste peti la agnoskon fare de la Respubliko de San Marino.

(Unuanime akceptita)

(Außerhalb der redaktionellen Zuständigkeit)

Richtlinien für die Manuskriptabfassung

Artikel von mehr als 12 Druckseiten Umfang (ca. 36.000 Anschläge) können in der Regel nicht angenommen werden; bevorzugt werden Beiträge von maximal 8 Druckseiten Länge. Außer deutschsprachigen Texten erscheinen ab 1982 regelmäßig auch Artikel in den drei Kongreßsprachen der Association Internationale de Cybernétique, also in Englisch, Französisch und Internacia Lingvo. Die verwendete Literatur ist, nach Autorennamen alphabetisch geordnet, in einem Schrifttumsverzeichnis am Schluß des Beitrags zusammenzustellen - verschiedene Werke desselben Autors chronologisch geordnet, bei Arbeiten aus demselben Jahr nach Zufügung von „a", „b" usf.. Die Vornamen der Autoren sind mindestens abgekürzt zu nennen. Bei selbständigen Veröffentlichungen sind anschließend nacheinander Titel (evt. mit zugefügter Übersetzung, falls er nicht in einer der Sprachen dieser Zeitschrift steht), Erscheinungsort und -jahr, womöglich auch Verlag, anzugeben. Zeitschriftenbeiträge werden nach dem Titel vermerkt durch Name der Zeitschrift, Band, Seiten und Jahr. - Im Text selbst soll grundsätzlich durch Nennung des Autorennamens und des Erscheinungsjahrs (evt. mit dem Zusatz „a" etc.) zitiert werden. - Bilder (die möglichst als Druckvorlagen beizufügen sind) einschl. Tabellen sind als „Bild 1" usf. zu numerieren und nur so zu erwähnen, nicht durch Wendungen wie „vgl. folgendes (nebenstehendes) Bild". - Bei Formeln sind die Variablen und die richtige Stellung kleiner Zusatzzeichen (z.B. Indices) zu kennzeichnen.

Ein Knapptext (500 - 1.500 Anschläge einschl. Titelübersetzung) ist in mindestens einer der drei anderen Sprachen der GrKG/Humankybernetik beizufügen.

Im Interesse erträglicher Redaktions- und Produktionskosten bei Wahrung einer guten typographischen und stilistischen Qualität ist von Fußnoten, unnötigen Wiederholungen von Variablensymbolen und übermäßig vielen oder typographisch unnötig komplizierten Formeln (soweit sie nicht als druckfertige Bilder geliefert werden) abzusehen, und die englische oder französische Sprache für Originalarbeiten in der Regel nur von „native speakers" dieser Sprachen zu benutzen.

*

Direktivoj por la pretigo de manuskriptoj

Artikoloj, kies amplekso superas 12 prespaĝojn (ĉ. 36.000 tajpsignojn) normale ne estas akceptataj; preferataj estas artikoloj maksimume 8 prespaĝojn ampleksaj. Krom germanlingvaj tekstoj aperadas de 1982 ankaŭ artikoloj en la tri kongreslingvoj de l'Association Internationale de Cybernétique, t.e. en la angla, franca kaj Internacia lingvoj.

La uzita literaturo estu surlistigita je la fino de la teksto laŭ aŭtornomoj ordigita alfabete; plurajn publikaĵojn de la sama aŭtoro bv. surlistigi en kronologia ordo, en kazo de samjareco aldoninte „a", „b" ktp.. La nompartoj ne ĉefaj estu almenaŭ mallongigite aldonitaj. De disaj publikaĵoj estu - poste - indikitaj laŭvice la titolo (evt. kun traduko, se ĝi ne estas en unu el la lingvoj de ĉi tiu revuo), la loko kaj jaro de la apero, kaj laŭeble la eldonejo. Artikoloj en revuoj ktp. estu registritaj post la titolo per la nomo de la revuo, volumo, paĝoj kaj jaro. - En la teksto mem bv. citi pere de la aŭtornomo kaj la aperjaro (evt. aldoninte „a" ktp.). - Bildojn (laŭeble presprete aldonendajn!) inkl. tabelojn bv. numeri per „bildo 1" ktp. kaj mencii ilin nur tiel, neniam per teksteroj kiel „vd. la jenan (apudan) bildon". - En formuloj bv. indiki la variablojn kaj la ĝustan pozicion de etliteraj aldonsignoj (ekz. indicoj).

Bv. aldoni resumon (500 -1.500 tajpsignojn inkluzive tradukon de la titolo) en unu el la tri aliaj lingvoj de GrKG/Humankybernetik.

Por ke la kostoj de la redaktado kaj produktado restu raciaj kaj tamen la revuo grafike kaj stile bonkvalita, piednotoj, nenecesaj ripetoj de simboloj por variabloj kaj tro abundaj, tipografie nenecese komplikaj formuloj (se ne temas pri prespretaj bildoj) estas evitendaj, kaj artikoloj en la angla aŭ franca lingvoj normale verkendaj de denaskaj parolantoj de tiuj ĉi lingvoj.

*

Regulations concerning the preparation of manuscripts

Articles occupying more than 12 printed pages (ca. 36,000 type-strokes) will not normally be accepted; a maximum of 8 printed pages is preferable. From 1982 onwards articles in the three working-languages of the Association Internationale de Cybernétique, namely English, French and Internacia Lingvo will appear in addition to those in German. Literature quoted should be listed at the end of the article in alphabetical order of authors' names. Various works by the same author should appear in chronological order of publication. Several items appearing in the same year should be differentiated by the addition of the letters "a", "b", etc. Given names of authors, (abbreviated if necessary, should be indicated. Works by a single author should be named along with place and year of publication and publisher if known. If articles appearing in journals are quoted, the name, volume, year and page-number should be indicated. Titles in languages other than those of this journal should be accompanied by a translation into one of these if possible. - Quotations within articles must name the author and the year of publication (with an additional letter of the alphabet if necessary). - Illustrations (fit for printing if possible) should be numbered "figure 1", "figure 2", etc. They should be referred to as such in the text and not as, say, "the following figure". - Any variables or indices occuring in mathematical formulae should be properly indicated as such.

A resumee (500 - 1,500 type-strokes including translation of title) in at least one of the other languages of publication should also be submitted.

To keep editing and printing costs at a tolerable level while maintaining a suitable typographic quality, we request you to avoid footnotes, unnecessary repetition of variable-symbols or typographically complicated formulae (these may of course be submitted in a state suitable for printing). Non-native-speakers of English or French should, as far as possible, avoid submitting contributions in these two languages.

*

Forme des manuscrits

D'une manière générale les manuscrits comportant plus de 12 pages imprimées ne peuvent être acceptés. Les références littéraires doivent faire l'objet d'une bibliographie alphabétique en fin d'article. Plusieurs oeuvres du même auteur peuvent être énumerées par ordre chronologique. Le prénom de chaque auteur doit être mentionné, au moins en abrégé. Indiquez le titre, le lieu et l'année de publication, et, si possible, l'éditeur des livres, ou, en cas d'articles de revue, le nom de la revue, le titre, les pages (p. ex. p. 317-324) et l'année dans cet ordre. On peut mentionner le titre des articles ayant fait l'objet de publications. Les publications d'un auteur parues la même année feront l'objet d'une classification (telle que a, b etc.). On citera dans le texte le nom de l'auteur, suivi de l'année de l'édition (éventuellement complété par "a" etc.). Evitez les notes en bas de pages.